GUK-H00683-25-P6154

Brit. J. Phil. Sci. 38 (1987), 141–158 *Printed in Great Britain* 141

501.05
B 774

25

Explanation, Independence and Realism in Mathematics

by MICHAEL D. RESNIK *and* DAVID KUSHNER

1 *Introduction*
2 *An Overview of Steiner's Theory*
3 *Some Details Concerning Explanatory Proofs*
4 *Mathematical Explanations*
5 *Independence Again*

1 INTRODUCTION

The interconnections between the various branches of mathematics, which are often striking and fascinating, readily provoke the philosophical imagination. Recently, Mark Steiner ([1978] and [1983]) has tried to base a case for mathematical realism on Euler's equation

$$e^{\pi i}+1 = 0.$$

P 6154

This elegant little identity connects a fundamental constant of analysis, *e*, with equally fundamental constants of arithmetic, algebra and geometry. It would be hard for a philosopher to pass it by.

Steiner distinguishes mere existence (in Quine's sense of being the value of a variable) from two types of reality—ontic and epistemic reality. We will concern ourselves here with the epistemic reality of mathematical objects, for Steiner not only appeals to Euler's equation in connection with their epistemic reality, he also devotes most of his discussion to that topic. Steiner [1983] proposes that an object is real in the epistemic sense in case it has independent descriptions. In mathematics this means that there are two different descriptions of the object and a proof that they are coreferential, but no *explanatory proof* that they are. On Steiner's view, *pi* is real in the epistemic sense because Euler's identity shows that *pi*, first described *geometrically* as the ratio of the circumference of a circle to its diameter, answers to the *analytic* description of *arg*(−1), where

$$arg(x) = y \quad \text{if and only if} \quad e^{yi} = -1.$$

Furthermore, these are independent descriptions, since Euler's identity is 'established by magic, as it were. No explanatory proof has been given' ([1983], p. 378).

We share Steiner's fascination with examples such as Euler's equation. We agree with Steiner about the *phenomena* of mathematical practice on

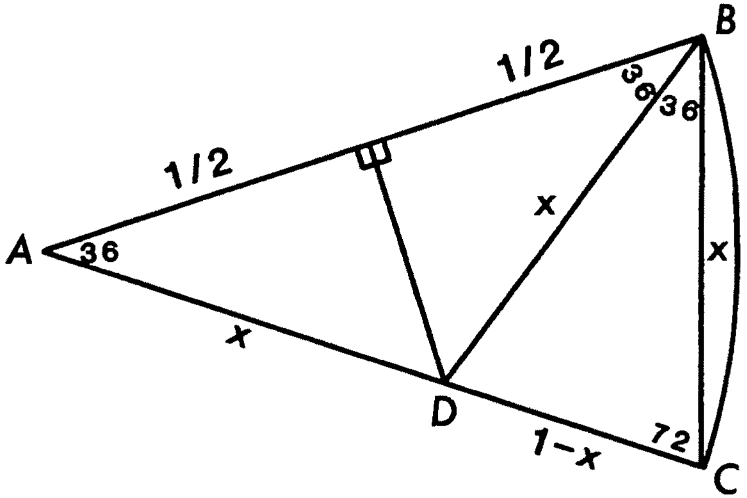

Fig. 1

which he focuses: mathematicians do request and offer explanations of purely mathematical matters; some proofs are very illuminating while others are rather opaque; some mathematical objects do appear in quite different and surprising contexts. However, we have serious questions to raise concerning Steiner's notion of explanatory proof and his use of it to ground the epistemic reality of mathematical entities. In the course of doing this we also hope to further illuminate the questions of explanation and epistemic independence in mathematics.

Before we proceed to discuss Steiner's views we would like to present another striking example of the interconnections between geometry and analysis. Euler's identity turns upon the power series expansions for the sine, cosine and exponential functions. A detailed consideration of it threatens to lead us into recondite areas of mathematical analysis. The example to follow should be accessible to those who remember just a modicum of geometry and algebra.[1]

Assume that we are given a circle of unit radius with a regular decagon inscribed in it. We will determine the length of the sides of this decagon. Let x be the length in question. Now x and the two radii it joins form an isosceles triangle whose apex angle is 36°. Thus the two base angles are each 72°. So if we bisect one with a line running to the opposite radius, we will divide our triangle into two isosceles triangles, *ABD* and *BCD* (see Figure 1). Since *BCD* is isosceles, the length of its side *BD* is x. By the same token, the length of *AD* is x, too, so the length of *DC* is $1-x$.

---

[1] We have pieced this example together from Courant and Robbins [1978], p. 123 and p. 302. For Euler's equation, see pp. 477–478.

Furthermore, triangle ABC is similar to BCD, so $x/(1-x) = 1/x$. This yields the quadratic equation

$$x^2+x-1 = 0, \tag{1}$$

whose positive root is $(1/2)(\sqrt{5}-1)$. We can also determine x by dividing ABD into two right triangles. Then we find that x equals $1/2 \cos 36°$. (For future reference we will refer to x *qua* geometric length as 'the inverse Golden Mean'. We name x thus because the Golden Mean equals $1/x$ or $(1/2)(\sqrt{5}+1)$.)

Now return to (1). This yields

$$x = 1/(1+x). \tag{2}$$

If we now substitute the right side for 'x' in (2), we obtain

$$x = 1/[1+1/(1+x)]. \tag{3}$$

Continuing in this way we can derive

$$x = 1/[1+1/[1+1/(1+x)]], \tag{4}$$

and so on. But the fractions on the right sides of these equations, each of which is identical to x, converge to the infinitely continued fraction

$$1/[1+1/[1+1/[1+[1+ \ldots .$$

Thus $(1/2)(\sqrt{5}-1) = x = 1/[1+1/[1+1/[1+[1+ \ldots = 1/2 \cos 36°$.

We have now described x geometrically (as the inverse Golden Mean) and analytically (as the value of a continued fraction). Does this mean that x is real in the epistemic sense? According to Steiner that depends on whether our proof is explanatory or not. Is it? Try to judge for yourself after you have read our exposition of Steiner's account of explanatory proofs.

2 AN OVERVIEW OF STEINER'S THEORY

We begin by reviewing Steiner's position. Steiner [1983] starts by endorsing Quine's argument that mathematical objects exist because quantifying over them is indispensible in contemporary science. But this leads him to raise the question of whether mathematical objects are real in the sense of 'existing independently'. Towards clarifying this Steiner distinguishes ontic and epistemic independence, and thus two types of reality. We will skip his discussion of ontic reality, since his main concern is with epistemic reality.

Roughly, Steiner's idea is that an object is real in the epistemic sense in case some of our epistemic access to it is conceptually independent of the theory which initially postulated it. Neutrinos are a good example. They were first postulated to account for energy missing from certain quantum interactions, and thus at first known only through this means. Later, however, they were detected via experiments that make sense independently of the specific laws that led to their postulation.[1] Drawing upon discussions by physicists and philosophers of science, Steiner concludes that 'what is real [in the epistemic sense] is what is susceptible to independent description' ([1983], p. 369). However, scientific reduction may 'unify' previously independent descriptions, thereby undermining the reality of the things described. In this connection Steiner remarks that there was a time when one might have argued for the epistemic reality of the electric field on the grounds that it can be independently described in terms of electrical theory and optics. But now that light has been identified with electromagnetism this option has been extinguished. 'Since the major goal of modern science is unification it follows that the goal of science is to show that nothing is real!' ([1983]), p. 370).[2]

Examples such as Euler's equation, the recurrence of the Golden Mean in geometry, difference equations and continued fractions, and the equivalence of the axiom of choice to principles in topology, algebra and mathematical logic strikingly suggest the epistemic reality of the mathematical realm. Although Steiner wanted to capitalize on this, he saw an immediate problem with speaking of independent descriptions in mathematics. To show that two mathematical descriptions refer to the same thing, we must *prove* their equivalence. But once we have done that how can we claim that they are independent?

This is where Steiner enlisted his previous work on mathematical explanation. Here he aimed to 'exploit the idea that to explain the behavior of an entity, one deduces the behavior from the essence or nature of the entity' ([1978], p. 143). Since mathematics holds in all possible worlds, appealing to essences outright will not do. Thus Steiner introduced the idea of a *characterizing property*, that is, 'a property unique to a given entity or structure within a *family* or domain of such entities or structures' ([1978], p. 143). Using this he then defined an *explanatory proof* as one which 'makes reference to a characterizing property of an entity or structure mentioned in the theorem, such as that from the proof it is evident that the result depends on the property' ([1978], p. 143). An explanatory proof that two descriptions are coextensive will, in effect, reveal that they are *connected* through essences, whereas a nonexplanatory proof will point to no such connection and leave the fact of their coextensiveness a 'mystery'.

---

[1] See also Salmon [1984], pp. 214–224 for a discussion of how the independent determinations of Avogadro's number convinced the scientific community of the reality of atoms.

[2] Steiner's notion of ontic reality seems to come to this: an object is real in the ontic sense just in case reference to it is ineliminable by paraphrase from the language of science. If this is so, then no mathematical objects are real in the ontic sense, for reference to numbers and functions may be eliminated in favor of reference to sets, but reference to sets in turn may be eliminated in favor of reference to functions (or even to ordinal numbers if one accepts $V = L$). On the other hand, no one has succeeded in eliminating all reference to mathematical objects. On Steiner's view they exist, but there is no fact to the matter of their ontic reality. Put this together with Steiner's remark that science aims to show that nothing is real in the epistemic sense and we are left with an uncomfortably tenuous grip on reality—in either sense.

3 SOME DETAILS CONCERNING EXPLANATORY PROOFS

Let us take a closer look at Steiner's account of explanatory proofs. His discussions and examples make it clear that an explanatory proof depends upon a characterizing property in the sense that (a) the proof fails to go through if we substitute another entity in the family which does not have the property and (b) by suitably 'deforming' the proof while holding 'the proof-idea' constant, we can get a proof of a related theorem ([1978], p. 143, p. 147). But, perhaps, an illustrative example will make matters plainer.

Here is one of Steiner's. Every positive integer can be factored uniquely into primes. This result is used in the following explanatory proof that for no positive integers a and b do we have $a^2 = 2b^2$—the crucial step in proving the irrationality of the square root of 2. Think of a^2 and $2b^2$ as written in their prime factorization forms. If the equation is true then there will be the same number of 2's on both sides. But there will be an even number of occurrences of 2 in the prime factorizations of each of a^2 and b^2, since they are squares. That means that there are an odd number of 2's on the right side of the equation and an even number on the left side—a contradiction.

It is easy to see that we can use this proof-idea to get new theorems. Substitute 3, 5, 7, or any other prime for 2 and the reasoning goes through without change—except we count the occurrences of the new prime rather than 2. Substitute 6, 8, 10 and you need to add a step: $6 = 2\times 3$, $8 = 2\times 4$, $10 = 2\times 5$, so in each case we will again have an odd number of 2's. In fact the only way to avoid having an odd number of occurrences of some factor is for the number substituted in the place of 2 to be a square. Thus we arrive at the general theorem: for positive intergers n, a and b, $a^2 = nb^2$ only if n is a perfect square; so the square root of n is either an integer or an irrational. (Did the proof make all this *evident*? Steiner apparently thinks so ([1978], p. 144).)

To see that this fits Steiner's criterion of an explanatory proof we must identify the characterizing property or properties used. Since Steiner remarks that the proof uses the prime factorization of 2, which uniquely characterizes it, and since we have seen how the proof must change as we substitute other numbers for 2 while holding the proof-idea constant, the characterizing property we seek must be *having 2 as its prime factorization*. When we replace 2 in deforming the proof, we must also replace this property by another of the form *having* x *as its prime factorization*. (Again, did the proof make this *evident*?)

How about a non-explanatory proof? Steiner gives few simple examples. One is the proof by induction of the identity

$$1+2+3+\cdots+n = n(n+1)/2.$$

Here we use this very equation to show that

$$1+2+3+\cdots+n+(n+1)=(n+1)(n+2)/2$$

but nothing about the proof shows us what happens when we substitute the sum of, say, the first n odd numbers or the first n even numbers. Another example is the (so-called) Pythagorean proof that

$$a^2=2b^2$$

is impossible if a and b are positive integers. It goes as follows: We can assume that a and b are relatively prime. Then if the equation is true, a^2 must be divisible by 2. But then (by a crucial lemma) so must be a. Thus we obtain a new equation,

$$(2a')^2=2b^2.$$

Dividing by 2 shows that b^2 and thus b must be divisible by 2, contradicting our assumption that a and b are relatively prime, Q.E.D.

According to Steiner this proof fails to be explanatory because it depends upon the crucial lemma that a^2 is divisible by 2 only if a is. We can generalize this and show that similar lemmas hold for other numbers but 'the proofs become more and more complex (where $p=5$, for example, one must square $5q+1$, $5q+2$, $5q+3$, and $5q+4$ and show that in no case is the result divisible by 5)' ([1978], p. 138).

We do not accept Steiner's account of explanatory proofs. We have doubts that any proofs *explain*, but more of that later. For now we wish to propose a proof that meets Steiner's criterion but doesn't explain and one which ought to explain if any proof does but fails to meet Steiner's criterion.

We do this with some trepidation, however. Steiner's criterion is framed in terms of the undefined ideas of a family of mathematical entities and of proof deformation. In applying his criterion we may encounter difficulties in determining family lines or the limits to proof deformation. Then there is the matter of the place of characterizing properties in the proof. Steiner states that an explanatory proof must *refer* to a characterizing property and make it *evident* that the result depends upon the property. Whether or not something is evident from a proof is relative to subgroups of the mathematical community, at best. A proof that explains to a mathematical logician may be anything but evident to a topologist. Then there is the matter of the explicitness of the reference to a characterizing property. That is a matter of style. None of the versions of the prime factorization proof of the irrationality of the square root of 2 that we have seen explicity refer to 2 being its own prime factor. But this is the property that we must vary as we deform the proof to prove the irrationality of other numbers. We also find it remarkable that a mathematician as eminent and philosophically reflective as Solomon Feferman proposed an intuitively explanatory proof to Steiner with the challenge to find the characterizing property used. To his credit, Steiner rose to the challenge, deformed the proof and arrived at another theorem ([1978], p. 151). But if the proof did not make the characterizing property evident to Feferman, to whom should it have made it evident? Finally, there is the question of defining the boundaries of a proof. Does the explanatory proof include the proof of the prime factorization theorem? If so, must the proof of the latter be explanatory too? The proofs of the prime factorization theorem that we have seen don't appear explanatory to us, so we suspect that Steiner would answer our questions in the negative. If so, how can explanatory proofs based upon nonexplanatory lemmas yield connections between 'essences'?

Let us return now to Steiner's nonexplanatory proof of the irrationality of the square root of 2. It depends upon remarking that if $a^2 = 2b^2$ then 2 divides a^2. Now the same is true for any positive integer that we substitute for 2. Furthermore, all but squares divide a^2 only if they divide a. Thus by deforming the proof as given so far, we obtain the same results that Steiner notes in favor of the proof based on prime factorization. Can we cite a characterizing property? How about this: being the least integer x such that any integer that x divides is also divisible by 2. Not very elegant, but it does single out 2 from the other positive integers and varying it appropriately yields the other results.

Yet Steiner still doesn't think that this proof meets his criterion because in each case we must reprove the crucial lemma and the proofs get 'more and more complex'. But here is a proof that removes this impediment.

Lemma: If $a = kn+i$ where $0 < i < k$ (*i.e.*, k does not divide a) and k is not a perfect square, then k does not divide a^2. Proof: By course of values induction on a. We know that

$$a^2 = (kn)^2 + 2kni + i^2.$$

So k divides a^2 only if it divides i^2. But $k \neq i^2$, since k is not square; so, applying the inductive hypothesis to i, k divides i^2 only if it divides i. But the latter is impossible since $i < k$.

To us this proof seems no less illuminating than the proofs we know of the prime factorization theorem. Thus pending further clarification of the question of where to draw the boundaries of a proof for the purposes of settling its explanatory value, we offer the Pythagorean proof of the irrationality of the square root of 2 as an example of a proof which meets Steiner's criteria but which fails, in his judgment, to explain.

Now for some proofs that would seem to qualify as explanatory if any do but don't appear to satisfy Steiner's criterion. Remembering the fate of Feferman's example, we stand prepared to be corrected.

Our first proof concerns the intermediate value theorem, which is one of the fundamental theorems of analysis. We shall prove the version of the theorem that states that if a real valued function f is continuous on the closed real interval $[a,b]$ and if $f(a) < c < f(b)$, then there is an x in $[a,b]$ for

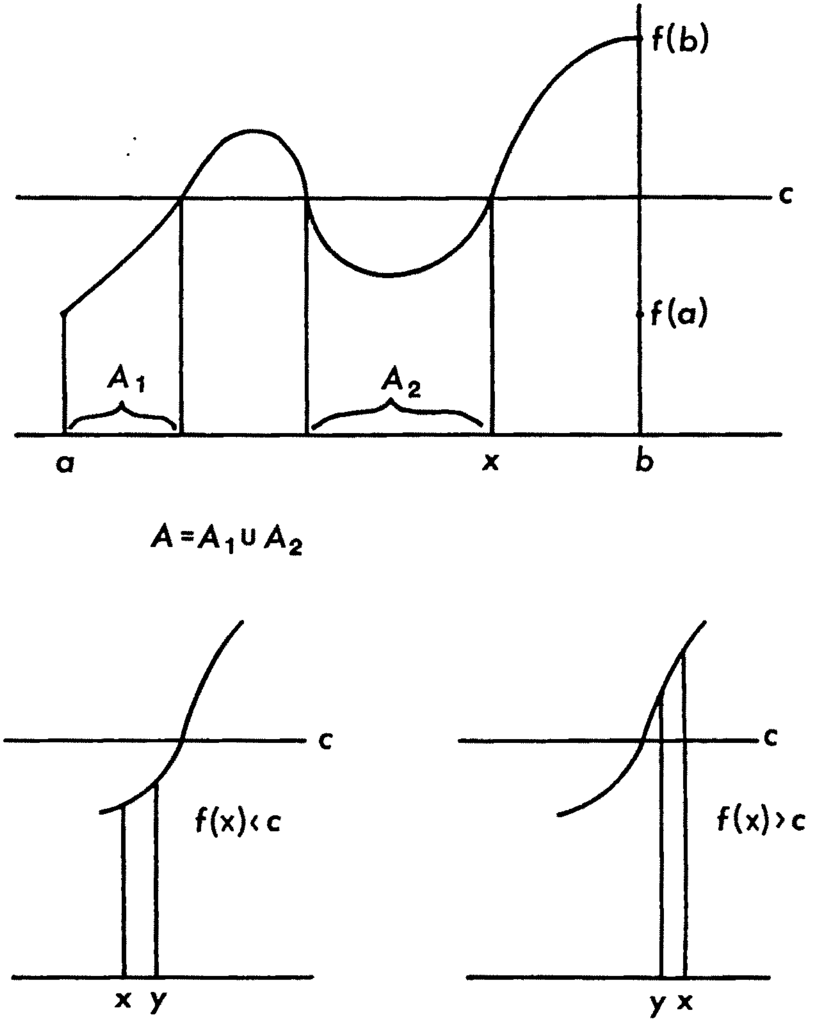

Fig. 2

which $f(x) = c$. We have modified a proof from Rudin [1953] to fit the illustrations in Figure 2.

Consider the set A of all the t in $[a,b]$ for which $f(t) < c$. A contains a and is bounded above by b; so, by the continuity of the real line, A has a least upper bound x with $x \leqslant b$. (The point x not only isolates the points of $[a,b]$ which yield values of f below c to its left but also is the leftmost point to do so.)

Since f is continuous on $[a,b]$, $f(x)$ is defined. We will prove by contradiction that $f(x) = c$. Suppose that $f(x) < c$. Because f is continuous, we know that we can pick a point y to the *right* of x (though perhaps very close to it) for which we have $f(y) < c$. This contradicts the fact that we have isolated all such points to the left of x. On the other hand, if $f(x) > c$ then for each point t in $[a,b]$ to the right of x we have $f(t) > c$. But again by the continuity of f we can find a point y' to the *left* of x such that for all t in $[y',b]$, $f(t) > c$. This contradicts the fact that x is the leftmost point to isolate the points of $[a,b]$ with values less than c.

We find it hard to see how someone could understand this proof and yet ask *why* the theorem is true (or what makes it true). The proof not only demonstrates how each element of the theorem is necessary to the validity of the proof but also what role each feature of the function and the interval plays in 'making' the theorem true. Moreover, it is easy to see that the theorem fails to hold if we drop any of its conditions.

Although the theorem has trivial expansions and much more abstract versions, which are well-known to students of analysis and topology, neither the theorem nor our proof is known to be 'deformable' to yield genuinely new results. In addition, as clear as the proof is, we find it hard to identify the characterizing properties on which it depends. According to Steiner, they should characterize something referred to in the theorem. To what does the theorem refer? Intervals, functions continuous on them and real numbers. The proof clearly depends upon properties of each—*e.g.*, that the function be continuous on an interval—but none of these come close to *characterizing* any particular function, interval or real number. Perhaps the theorem is really about the class of real valued functions over the real numbers drawn from the family of all classes of functions. This analysis is favored by the use of the same proof-idea in intermediate value theorems about continuous functions on domains other than real intervals. Yet for the proof to count as explanatory in Steiner's sense it must make plain how the theorem changes to new theorems as we move from the class of continuous functions over real closed intervals to any other class of functions in the family. But the proof doesn't tell us anything about discontinuous functions. Perhaps, then, we should restrict the family to classes of functions continuous on connected sets. But then we simply get the intermediate value theorem rephrased to fit the cases at hand. We don't get anything like the interestingly new theorems that pepper Steiner's discussion.

Another example is the Henkin proof of the completeness of first order logic. The proof proceeds by showing that any set of sentences that is deductively consistent (in the sense of first order logic) can be extended to a maximally deductively consistent set which can then be used to specify a model for the original set. The proof is generally regarded as really showing what goes on in the completeness theorem and the proof-idea has been used again and again in obtaining results about other logical systems. Yet again it is not easy to identify the characterizing properties on which it depends.

One problem is that all the versions of the theorem we have seen are ostensibly about particular formal systems. They state that this or that formal system is complete. Yet is is clear, and the proof makes it clear, that the proof works for other formal systems that codify the same set of deductions. So the proof cannot turn on a characterizing property of the particular formal system for which it happens to be stated. By analyzing the proof we could find a set D of deductions which are sufficient for the first-order completeness of a formal system. Thus, perhaps, the theorem is really about (the class of) those formal systems which are first-order sound and generate the set D. But does the proof make it plain how the theorem will fail when we move to formal systems which lack this property? Despite the wide use of the Henkin proof-idea, we think not. First, we think (though in correspondence Steiner disagrees) that the proof does not make it clear that when we apply the proof-idea to second-order logic, we must change the sense of model to allow for non-standard models, nor that when we apply it to modal logics, we must use many maximally consistent sets, *etc*. Second, if the proof did show this then we must have the wrong characterizing property, since in the second-order case we must change the notion of model as well as the set of deductions. Finally, we don't get related theorems when we try to apply the proof-idea to incomplete systems.

Let us try regarding the theorem as really about first-order consequence—showing that it has a complete proof procedure. If we do this, then we can take the relevant family as the set of consequence relations and characterizing properties will be readily at hand in the form of the conditions defining the consequence relations. However, we still have problems about the deformability of the result into related results where completeness fails. For example, does the Henkin proof yield a related theorem about number theoretic consequence? That depends on how far we can deform a proof before destroying the proof-idea: The compactness theorem falls out of the Henkin proof. This can be used to construct non-standard models for any consistent first-order number theoretic system, showing that such systems cannot be complete for number theoretic consequence.

At the end of [1978] Steiner remarks that to account for explanatory proofs in contemporary mathematics—he cites Galois theory and algebraic topology—he must weaken his criterion by admitting proofs which turn on partial characterizations. This would answer some of the questions we raised about characterizing properties, since it is clear which partial characterizations are involved in the proofs we have cited. However, this still will not deal with our questions about deformability.

Furthermore, once partial characterizations have been admitted we lose our grip on the idea that explanatory proofs reason about essences. Since Steiner describes a characterizing property simply as a property that singles out some object in a family of objects, a property that only partially characterizes is presumably just some property of an object in a family. Of course, an explanatory proof must not only use a property, it must also be deformable into related results by substituting other properties. But we find it hard to see how this makes essences of the properties used or how this conception of explanatory proof can serve as a satisfactory basis for clarifying the notion of independent mathematical descriptions.

4 MATHEMATICAL EXPLANATIONS

Having encountered a number of questions concerning Steiner's account, we thought it worth making a fresh start on the question of mathematical explanation and turned to the philosophy of science for insights. Unfortunately, we found no topic more hotly debated than the nature of scientific explanation. Furthermore, given the views of some leading philosophers of science, there can be no mathematical explanations. For, according to them, we explain in terms of causal processes, and these are conspicuously absent from the realm of mathematics. These controversies struck us as all the more ironic because, although from Aristotle onwards empirical science has acknowledged the production of explanations as one of its major goals and accomplishments, this is not an acknowledged goal of mathematical research. Mathematicians rarely describe themselves as explaining, and Steiner's work is one of the few philosophical accounts of mathematical explanation. Given such evidence that the *practice* of explaining mathematical phenomena has been barely acknowledged, one could hardly expect that testing descriptive or normative accounts of it would be an easy task.

Our attempts to evaluate Steiner's views on proofs confirmed this suspicion, as we have already seen. Moreover, even Steiner's clear cases can be debated. Take the two proofs of the irrationality of the square root of 2, which we discussed in the last section. Like Steiner, Davis and Hersh [1982] tout the prime factorization proof as a definite advance—but not on the grounds that it *explains*. Rather it 'exhibits a higher degree of aesthetic delight' (p. 299). Why? Because the proof 'seems to reveal the heart of the matter, while [the Pythagorean proof] conceals it, starting with a false hypothesis and ending with a contradiction' (p. 299). (But note that *both* proofs start 'with a false hypothesis' and end with a contradiction!)

Despite these pessimistic findings, there are a few things that can be said about explanations in mathematics. First, if to explain something is to give a systematic account of it, then mathematicians do explain: many of the greatest achievements in mathematics, such as Euclid's axiomatization of geometry or the Newton–Leibniz calculus, are systematic accounts of previously scattered results. Steiner's example of Chang and Keisler's stated aim of explaining preservation phenomena in model theory fits this view of mathematical explanation (Steiner [1978], p. 135). On the other hand, explanation in the sense of systematization does not entail the existence of explanatory proofs: according to Steiner, it's Polya's (rather than Euclid's) proof of the Pythagorean theorem that is explanatory ([1978], p. 138, p. 144; [1983], p. 377).

Second, it is certainly appropriate to ask why-questions concerning mathematical entities and results—these are at least symptomatic of requests for explanation. Moreover, many of the answers to such questions closely resemble the answers to requests for scientific explanations. Thus by citing the fact that only a right triangle can be decomposed without remainder into two similar triangles, we can explain (at least in the sense of 'making it clear') why the Pythagorean theorem holds for only right triangles. This favors Steiner's view because Polya's proof turns on this feature of right triangles. But not all requests for explanation in mathematics are answered by proofs. It seems plausible to us, for example, to regard the definition of addition in terms of unions as explaining its commutativity and associativity. And we might explain why the intermediate value theorem must be restricted to intervals by noting that the only proper subsets of the real line which are connected (in the topological sense) are intervals. This could point us in the direction of counter-examples (and thus proofs) that show that the theorem fails in the case of, say, the disjoint union of a finite number of intervals (because such a set is not connected), but such counter-examples do not look like Steiner's explanatory proofs to us.

Third, mathematical explanations often point to a number of results or informal glosses of them rather than a single theorem. Why is second-order arithmetic categorical while first-order arithmetic is not? Because with the former we can refer to all sets of numbers and require the induction axiom to be true of all of them, whereas with the latter we can only require the induction axiom to be true of at most a countable number of number sets, which opens the door to non-standard models. Why can a function be continuous at a point and fail to have a derivative there? Because the derivative of a function at a point indicates the slope of the graph of the function at that point and *at the points very close to it*; thus if a function changes continuously but rapidly at a point then its derivative may go undefined. (Example: moving from left to right the graph of the absolute value function slopes downward at a 45° angle until it hits the point (0,0) and then it starts upward at the same angle.)

None of this contradicts Steiner; he has already admitted that there are other explanations in mathematics besides those given by proofs. He aimed only to give an account of explanation by proof ([1978], p. 147). In view of the difficulties we have found with his account, we should ask whether there are explanatory proofs.

Now there is no doubt that there are proofs which, in the words of Davis and Hersh [1982], 'seem to reveal the heart of the matter' or show what is really going on, to use another phrase of conversational mathematics. So if any proofs explain, these qualify as candidates. But how can we decide whether they explain? We could try pitting intuitions (considered judgments) against each other, but our intuitions concerning candidate explanatory proofs are weak, sparse and ill-defined. We tried reflecting, for example, on the differences among various consistency proofs for number theory and on the distinction between the purely model theoretic proof of the Löwenheim–Skolem Theorem using the axiom of choice and the mixed proof which detours through the completeness theorem. We found that different proofs provided more information or different information, but none struck us as clearly explanatory.

The why-question analysis of explanation of Van Fraassen [1980] points the way towards a resolution. According to Van Fraassen, to request an explanation is to pose a why-question. But a why-question consists of three elements: its topic (Why is this so?), its contrast class (Why this one rather than one of those?), and a relevance relation which determines whether a given proposition counts as an answer to the question. The context in which a why-question is uttered determines its topic, contrast class and relevance relation, and thus its identity. For instance, in one context somebody who asks 'Why is truth-functional validity decidable?' might be asking why it is decidable rather than undecidable and might be prepared to find any decidability proof an acceptable answer. In another context a questioner using the same form of words might be asking why truth-functional rather than quantification validity is decidable and might be prepared to accept an answer that points out that a truth-functional schema, in contrast to a quantificational schema, has finitely many interpretations. In another context, however, that answer might not serve, for there are decision procedures for the validity of some kinds of quantificational schemata, despite the infinitude of their interpretations.

Without committing ourselves to the details of Van Fraassen's analysis nor its *prima facie* ontic commitment to propositions we shall adopt its moral that nothing is an explanation *simpliciter* but only relative to the context dependent why-question(s) that it answers. This accounts for our difficulty in assessing Steiner's examples and in identifying explanatory proofs on our own. Whether or not a given proof counts as an explanation depends upon the why-question with which it is approached. If you simply wanted to know why a result is true (rather than false) and were prepared to accept any proof as an answer then you would count all its proofs as explanatory. But you might want to know more. For instance, in addition to wanting to know why the Pythagorean theorem holds you might want to know why it holds *only* for right triangles. Then not every proof of the theorem will contain an answer for you.

This approach will also account for the impression that some proofs seem explanatory in their own right. As our previous examples have shown, one need not give a proof to answer a mathematical why-question. However, one must cite relevant mathematical results. Every mathematical proof presents information organized to provide convincing reasons for its conclusion. (Mathematicians do not discover proofs by simply blindly deducing conclusions from previous results; rather by first becoming thoroughly familiar with a mathematical structure, they are able to see what is true of it and how these truths depend upon the basic properties of the structure.) Now the so-called explanatory proofs, the ones which 'reveal the heart of the matter', present more information and do so more perspicuously than do 'nonexplanatory' proofs of the same results. Thus they provide the ingredients for answering more why-questions than other proofs. But they are not explanatory in and of themselves. Van Fraassen puts the point well: 'So scientific explanation is not (pure) science but an application of science. It is a use of science to satisfy certain of our desires; and these desires are quite specific in a specific context, but they are always for descriptive information. . . . Hence there can be no question of explanatory power as such . . .' ([1980], p. 156). (Our endorsement of this passage does not extend to Van Fraassen's anti-realist conclusions.)

We can also see how Steiner might have been led to develop an account of explanatory proofs in terms of characterizing properties. When a theorem is about a particular object in a family of mathematical objects, it is natural for us to ask why the theorem is true of that object and not the others, what it is about the object that is responsible for this and how the theorem changes as we turn to the other objects in the family. Steiner's deformable proofs are ideally suited to presenting the information to answer such questions.

Finally, we can account for what is probably the most basic intuition behind the idea that there must be explanatory proofs as such, namely, that all proofs convince us that the theorem proved is true but only some leave us wondering why it is true. We have this intuition, we submit, because we have observed that many proofs are perfectly satisfactory as proofs but present so little information concerning the underlying structure treated by the theorem that they leave many of our why-questions unanswered. In reflecting on this, we tend to conflate these unanswered why-questions under the one form of words 'why is this true?' and thus derive the mistaken idea that there is an objective distinction between explanatory and nonexplanatory proofs.

5 INDEPENDENCE AGAIN

Obviously Steiner's account of independent decriptions is in trouble if there are no explanatory proofs. However, his examples are so fascinating and his conclusions so suggestive that we are reluctant to dismiss either out of hand. Let us return to our example of x (or the inverse Golden Mean).

We described x geometrically (as the inverse Golden Mean) and analytically (as the value of a continued fraction). The connection here is certainly startling to those unfamiliar with it. But having rejected Steiner's conception of explanatory proof, can we draw any conclusions from it concerning the epistemic reality of x?

Just to focus our intuitions, note that since $\sin 54° = \cos 36°$, we may also describe x as $1/2 \sin 54°$. Now consider the following identities, each of which shows that one of the alternative descriptions of x is equivalent to another one:

(*a*) $\sin 54° = \cos 36°$
(*b*) the inverse Golden Mean $= 1/2 \cos 36°$
(*c*) $1/2 \cos 36° = 1/[1+1/[1+1/[1+[1+\cdots$.

Our task is to determine whether we can rehabilitate Steiner's notion by finding an epistemologically significant difference between (*a*) and (*b*), on the one hand, and (*c*) on the other.

Now one might be tempted to respond that there is a clear epistemological difference here, for (*a*) is an immediate consequence of the definitions of sine and cosine whereas (*c*) is a 'deep' result. But this will not do for two reasons: First, (*a*) is not just a definitional consequence, it also depends upon facts about the sum of the angles in a right triangle. Second, we are trying to put *both* (*a*) and (*b*) on the same side of an epistemological partition; although it is plausible that the concepts involved in (*b*) are part of the same conceptual network, (*b*) itself is no candidate for a definitional truth.

Here is a more plausible attack. There is a 'natural' formulation of geometry *cum* arithmetic in which the terms contained in both (*a*) and (*b*) can be defined whereas some of the terms contained in (*c*) can only be defined in a logically stronger theory. The same is true of Steiner's example, since *pi* can be defined in elementary geometry, but defining $arg(-1)$ requires a stronger theory. By making this idea more precise we might find the division Steiner seeks, since one might argue that the facts about definability just mentioned show that some of the descriptions of the inverse Golden Mean and *pi* draw upon concepts belonging to different conceptual networks.

But we will not attempt the exercise in precision, since we have no intention of defending a proposal along these lines. A major problem with the proposal is that it ties the notion of epistemic reality to the particular *formulations* of mathematical theories that happen to hold sway in current mathematical practice. As a consequence, a simple reformulation of one of our current theories could cause previously independent descriptions to become dependent. Here is a simple artificial example of what we have in mind. Let $T1$ be the complete first-order theory of the less than relation over the natural numbers. Using the primitives of $T1$, '0' and '$<$', it is easy to define the successor relation within $T1$: Sxy just in case y is the least number greater than x. Accordingly, if $T1$ were our 'natural' theory of the natural number sequence, 0 would not be independently describable as *the least natural number* and as *the natural number which is not a successor*. Now let $T2$ be the complete first-order theory of the successor relation among the natural numbers. Its primitives are '0' and 'S', and '$<$' is not definable within it. To define '$<$' we must extend $T2$ to a second-order theory. Thus, if $T2$ were our 'natural' theory of the natural number sesquence, 0 would be independently describable as ***the least natural number*** and as ***the natural number that is not a successor***. At a deeper level there is a characterization of equi-cardinality within (ZF) set theory which, depending on the specific set of axioms and definitions for ZF used, varies from the status of a definition to a version of the deep Shroeder–Bernstein theorem to an equivalent of the axiom of choice. Furthermore, each set of axioms and definitions seems equally 'natural' (Quine [1971], sect 29).[1]

Steiner's approach to epistemic realism depends upon the divisibility of mathematics into conceptually independent domains. 'Holism has its limits,' he proclaims ([1983], p. 369). Now it is not clear that successfully distinguishing between explanatory and non-explanatory proofs would have drawn the lines as Steiner desired. But we need not worry about that; the notion of an explanatory proof is not viable. Nor should we trust the technical notions of definability or logical independence to serve Steiner's purposes, for, as we have seen, these must be relativized to particular theory formulations. While we cannot establish conclusively that holism in mathematics is correct, it seems to us that one need only reflect on such disciplines as analytic geometry, algebraic topology, algebraic geometry or the grand foundations such as set theory or category theory to see how implausible it is that fixed conceptual boundaries between mathematical disciplines exist. (Recall, too, that analysis has its historical and conceptual roots in geometry and that its arithmetization gave it 'independent' foundations.)

We will conclude by remarking on other aspects of mathematics which might be cited in support of its epistemic reality. One is the following version of Benacerraf's observation: the natural number sequence is isomorphic to infinitely many mathematical progressions, but there is no mathematical evidence to decide whether it is identical to any single one of them (Benacerraf [1983]). Now since there is no question of our being able to prove or refute the claim that the natural number sequence is one of the progressions in question, it is tempting to say that the natural number sequence is independently describable as this or that progression. Since Benacerraf's point applies to every mathematical structure, their epistemic reality would follow immediately.

Unfortunately, this will not do. There are serious technical problems with making sense of the idea that an ***identical structure*** answers to the various putative descriptions (Resnik [1981]). What is more relevant to our present concerns, however, is that the so-called independence of the descriptions in question is dependent upon a prior division of mathematics into separate theories. The claim that the natural number sequence is independently described by $T1$, $T2$, or by geometry or set theory presupposes that these theories are independent parts of mathematics. We are back to the holism issue again and to the problem that a mere reformulation of a branch of mathematics will shift the lines of division within mathematics.

---

[1] Quine [1976] uses similar observations about postulation and definition in a somewhat different connection.

The same difficulty confronts another tempting example, drawn this time from recursive function theory. In the early 1930s several researchers, working unaware of each other, proposed a number of mathematical explications of the informal notion of a mechanically computable function. Turing gave a characterization in terms of Turing machines, Gödel and Herbrand gave one in terms of systems of equations, Church did it with the lambda calculus, Post used his Post systems, and there were others. Yet it was soon proven that each proposal identified the same class of functions—now known as the recursive functions. This coincidence is often cited as evidence that we have found the mathematical equivalent of a natural kind.

However, there is little to the idea of independent definitions of the recursive functions besides historical independence. This is easily seen by examining the proofs that the various characterizations coincide. Roughly, each proof is to the effect that each derivation in a certain formal system is associated with a derivation in a certain other formal system. In short, these are theorems in proof theory. Now one might argue that proof theory and model theory are independent mathematical subjects, but it is surely stretching things to argue that the proof theory of Post systems is independent of the proof theory of, say, the lambda calculus. Again the holism issue must be faced.

This casts enough doubt on Steiner's approach to epistemic realism, we think. Euler's equation and similar examples are so striking that it is hard to shake the temptation to seek in them some philosophically significant feature of mathematics. We believe the foregoing shows that such an enterprise should be approached skeptically. But for now we have nothing else to add.[1]

---

[1] Or almost nothing else!

Some readers might have been struck by the similarity between Steiner's approach and the Putnam-Boyd arguments for scientific realism (on the grounds that realism is the only view that does not make a miracle of the success of science). Steiner criticizes Putnam for arguing that the only plausible explanation for the frequency with which unexpected interconnections crop up in mathematics is that mathematics describes something that is entirely independent of our ideas and language. We agree with Steiner that we need not appeal to such an independent reality to explain the appearance of Euler's equation and the like. Game formalists, for instance, could easily explain the phenomenon by arguing that the branches of mathematics are very complex games, whose features we cannot fully predict in advance—just as we can't fully predict the behavior of a complex computer program. On the other hand, since we tend to replicate many properties of extant games when we create new ones, it is no miracle to the game formalist that there are interconnections between the various branches of mathematics. Off hand, it seems to us that only intuitionism in its

We are grateful to Catherine Elgin, Susan Hale, William Lycan, David Resnik, Jay Rosenberg, Geoffry Sayre-McCord and Mark Steiner for their comments.

University of North Carolina at Chapel Hill
Princeton University

REFERENCES

BENACERAFF, P. [1983]: 'What Numbers Could Not Be,' in *Philosophy of Mathematics*, Second Edition, P. BENACERAFF and H. PUTNAM (*eds.*). Cambridge: Cambridge University Press.

COURANT, R. and H. ROBBINS [1978]: *What Is Mathematics?* Oxford: Oxford University Press.

DAVIS, P. J. and R. HERSH [1982]: *The Mathematical Experience*. Boston: Houghton Mifflin Co.

QUINE, W. V. [1971]: *Set Theory and Its Logic*, Second Edition. Cambridge, Mass.: Harvard University Press.

QUINE, W. V. [1976]: 'Carnap and Logical Truth,' in *The Ways of Paradox* by W. V. Quine. Cambridge, Mass.: Harvard University Press.

RESNIK, M. D. [1981]: 'Mathematics as a Science of Patterns: Ontology and Reference,' *Nous*, **15**, pp. 529–550.

RUDIN, W. [1953]: *Principles of Mathematical Analysis*. New York: McGraw-Hill Book Co.

SALMON, W. [1984]: *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.

STEINER, M. [1978]: 'Mathematical Explanation,' *Philosophical Studies*, **34**, pp. 135–151.

STEINER, M. [1983]: 'Mathematical Realism,' *Nous*, **17**, pp. 363–385.

VAN FRAASSEN, B. C. [1980]: *The Scientific Image*. Oxford: The Clarendon Press.

---

strictest and most subjective form and radical conventionalism will have difficulties providing plausible explanations for surprises in mathematics.

Catherine Elgin suggested to us that the interconnections between *historically* independent mathematical theories do provide good reasons for believing in the reality of mathematical objects—the holism issue notwithstanding. We agree to this much: The (historically) independent determinations of Avogadro's number provided good reasons for scientists to countenance atoms. Similarly, the (historically) independent characterizations of the recursive functions furnished strong evidence to mathematical logicians that they had successfully explicated the notion of an algorithmic process. In short, we agree that historically independent epistemic routes to the same objects or phenomena constitute good grounds for acknowledging the reality of objects or phenomena in question. If that is all there is to calling them 'real', then we have no quarrel.

But one might be tempted to distinguish objects that merely exist from those which both exist and are 'real'. One might further be tempted to infer that the existence of the former is more tentative than that of the latter. This could lead to a methodological principle to the effect that when we revise a theory we should preserve its 'real' objects but are free to excise the 'unreal' ones. We have no objection to this principle—so long as it is restated to eliminate the use of the word 'real' and its synonyms. But it should not be given absolute sway on existential decisions: on balance better science might be achieved in a particular instance by repudiating some 'real' objects while embracing some that merely exist.

Brit. J. Phil. Sci. 38 (1987), 159–179 *Printed in Great Britain*

The 'Economic' Approach to the Philosophy of Science

by GERARD RADNITZKY*

(1) What may be gained by applying concepts generalised from economics to methodological problems? The perspective of cost-benefit analysis ('CBA' for short) may help the researcher to see what sorts of questions he should take into account when dealing with particular methodological problems. This claim is supported by applying generalised CBA-thinking to two standard problems of methodology. (2) In the practice of research the handling of basic statements does not normally constitute any problem, and no conscious decision is involved. In the methodological reconstruction the key questions are: 'How can a particular basic statement be criticised?' and 'What are the costs of defending a statement that is "problematic"?' The problem of the empirical 'base' is an investment problem: whether or not to invest time and effort into processing a particular basic statement into a falsifying hypothesis for the theory we wish to test. The valuation of the costs of rejecting or, as the case may be, of defending a basic statement, are objective. (3) With respect to theories, in basic science, the issue is not one of acceptance or rejection of a single theory, but rather of theory preference. Both the rational response to a falsification and rational theory preference are governed by CBA-considerations. The option for one of two competing theories is based upon a CBA where the valuation of benefits and costs is objective. Theory change is an objective process, at least in those fields, where theorising is closely controlled by empirical testing: the better theory drives out the less good theory. The costs of defending a theory that is less good than its competitor are mainly epistemic resources forgone. The use of CBA in methodology not only is compatible with Popper's position, but it may pay to view Popper's methodology as an application of CBA to epistemic situations.

1 *Clarification of the Concepts of Economic Approach and of Methodology as they are Used in this Essay*
2 *Cost-Benefit Analysis Applied to the Problem of the Empirical Basis*
3 *Cost-Benefit Analysis Applied to the Problem of Rational Theory Preference*

Karl Popper stressed, justly, that all life is problem solving. Resources, scarcity, competition and selection, together with the laws of nature that govern them, apply to all organisms, not just to man (Hirshleifer [1985]; *cf.* also Ghiselin [1978], p. 233 and Ghiselin [1987]). Recently concepts

Received 24 June 1985.

---

* Part of the ideas presented in this paper are drawn from the author's earlier work in preparation for a workshop on the application of the economic approach to various fields ranging from biology to law, which took place in Vienna in 1984 and the results of which have been published in a collection edited by G. Radnitzky and P. Bernholz.

generalised from economics have been applied to fields of study that have traditionally been thought to lie not only beyond the proficiency of economics but even outside the applicability of the perspective of economic thinking. They have been applied successfully in the sense that this application has produced testable consequences and new insights. The field of study in which concepts generalised from economics have been applied include biology, psychology, ethnology, political science, and the evolution of law. When applied to human phenomena the 'economic' approach has enabled us to recognise that certain situations that we had come to regard as not market-like are in fact market-like in important respects. For example, in the light of the economic theory of politics we can conceptualise democracy as a competitive vote-producing and vote-trading system; in the light of Gary Becker's studies we can conceptualise the family as analogous to the firm, and such studies as Bernard Heinrich's *Bumblebee Economics* provide interesting as-if-rational explanations of animal behaviour. (For a sample of applications of the 'economic' approach outside the traditional areas of economics see Radnitzky and Bernholz [1987].)

To forestall possible misunderstandings I want to mention two things. First, the economic approach has nothing to do with explanations in terms of 'economic interest', or financial matters, of costs in monetary units, etc. A glance at the concept of cost can exemplify this. For every act there is a cost. Money is just *one* means of measuring costs: there is also the time invested and the benefits forgone when using resources in the act, there may be psychic costs such as pangs of conscience, political costs, and so forth. Second, 'economic approach' as conceived here has nothing to do with Marxist 'economic approach'. The latter is a doctrine based on the assumptions that the organisation of production determines the social and political structure, and that what matters are material goods and processes as well as the conflict between capitalists and workers. Predictions made with the help of these assumptions led to falsifications; these in turn evoked immunisation strategies, which transformed the Marxist theory into a pseudo-scientific doctrine.

In this essay I wish *to investigate what may be gained from applying the 'economic' approach, in particular cost-benefit thinking, to the methodology of research.* Is it more than *une façon de parler* and a heuristic device? I conjecture that applying cost-benefit analysis ('CBA' for short) in the philosophy of science will not solve any methodological problems, but that the CBA-frame may provide an organising schema that helps the researcher to see what sorts of questions he should take into account when dealing with such problems. This claim will be illustrated by applying CBA to two central problems of methodology: the problem of the empirical 'base' and the problem of rational theory preference. In the first case it helps to clarify the nature of the problem; in the second example it helps us to see what exactly the methodological problem is. Popperian methodology may be viewed as an application of the economic approach to epistemic situations.

1 CLARIFICATION OF THE CONCEPTS OF 'ECONOMIC' APPROACH AND OF METHODOLOGY AS THEY ARE USED IN THIS ESSAY

1.1 Although the *economic approach* is part of the Neo-classical tradition, probably most readers will associate it primarily to the Chicago School and in particular to the work of Gary Becker. Becker postulates maximising behaviour, market equilibrium and stable preference structure, a uniform model of man so to speak. Some of the exponents of Austrian economics have considered these postulates to be too strong. (For a comparison see, *e.g.*, Gray [1987].) In general, Austrian economics has been more concerned with the searching process than with the analysis of the final results of action, since they have stressed the fallibility of human knowledge and uncertainty. But they too view human action as characterised by the intention to attain a state of affairs preferred to that which exists. Action in the full sense is predicated of individuals (and by extension of households and firms), but not of societies. Action has to do with problem solving and thus also with knowledge problems. One central problem is the collocation of relevant knowledge and decision rights. Popper added the method of 'situational analysis'—or, as I would prefer to call it, 'the rational-problem-solving approach'—to the MS of *The Poverty of Historicism* in 1938 and he conceived it as 'an attempt *to generalise the method of economic theory (marginal utility theory) so as to become applicable to the other theoretical social sciences*' (Popper [1976], pp. 117 f., italics in the original). Both Friedrich von Hayek and Popper regard free competition in ideas and their continuous testing as the proper means to improving our knowledge. Hayek goes further than Popper and emphasises that free competition in action is the basic discovery procedure for the best way to pursue human aims.

In this context the expression 'economic approach' is used in the widest sense. It includes Popper's rational-problem-solving approach as well as as-if-rational explanations of biological phenomena. What matters here is that the parties mentioned above agree that the rationality principle—be it conceptualised as maximising or in the weaker form of Popper's situational logic—is applicable to all human action, even if that action occurs in situations that are not market-like.

Man is a chooser. All rational choices involve the weighing up of benefits and costs. Hence, CBA is the core of the economic approach/rational-problem-solving approach. Strictly speaking, what is compared are benefits and forgone benefits, *i.e.*, opportunity costs, where the forgone utilities of the next best alternative figure as final costs. In economics the various benefits and forgone benefits associated with competing capital investments are made commensurate with one another by converting them to monetary units. But, if we generalise the concepts of CBA, we recognise that not only in economic life but also in daily life we operate within the CBA-frame, most of the time. If this is a discovery, then it is of the same sort as that of

Monsieur Jourdain when he discovered that he had spoken prose all his life.

For about three decades, CBA has become associated with the public sector economy. This use of the CBA-frame is full of problems. The public sector decision-maker's objective functions is said to be the maximisation of "social welfare", which is a highly elusive concept. The value assigned to the sacrificed opportunities depends on what opportunity sets the particular decision makers take into account, and on how they evaluate the various expected benefits. Since these evaluations cannot be but *subjective*, opportunity costs for different individuals are not commensurable. Hence, a CBA in welfare economics that 'pretends to base its arguments on inter-personal comparison of ascertainable utilities, lacks all scientific foundation' (Hayek [1979], p. 201).

For *individual* actors and the *private sector economy* the application of the CBA is less problematic. For instance, for ascertaining *ex post*—with the benefit of hindsight—which of the investment projects that were in the individual's opportunity set at the moment of decision has so far been the "best", various tools are available such as, *e.g.*, the internal rate of profitability method (*cf.*, *e.g.*, Radnitzky [1964]). Although CBA is a conceptual instrument for optimising goal realisation and, hence, implicit in all purposive action, it can be applied meaningfully only if the preference structure of the decision maker is given and several aims compete for scarce resources. That means that CBA is not applicable in situations where one single value is set absolute, in "ultimate" choices. In such situation, if means are scarce, the problem is not an economic but a technological problem (Kirsch [1971], p. 70). It also follows that CBA is applicable only if the expected benefits of the projects to be compared can be expressed in such a way that there is a reference basis common to all, *i.e.*, if it is possible to compare the benefits to be gained from the project chosen and the benefits forgone. For an individual actor this requirement is fulfilled, because for an individual it is always possible to derive from his preference structure a criterion or measure that is intrasubjectively valid, *i.e.*, valid for him. He can even choose between "incommensurable" benefits, because for an individual there is always a *tertium comparationis* which may simply consist in the *subjective* evaluation that he prefers this to that. Hence, the insurmountable problems of intersubjectively meaningful comparison that beset the application of CBA in public sector economy need not arise for the individual decision maker.

When CBA is applied to the methodology of research, we have to distinguish between the situation of a particular, individual researcher and the methodology of research as a special discipline—a discipline which, perhaps, is even more imperialistic than the economic approach, since it deals with methods, criteria, argumentations, etc., in *any* scientific discipline. In the case of an individual researcher, CBA faces the same sorts of problems as it does in any individual decision-making process. Whether or not the utility function of a real person includes scientific progress as an aim is irrelevant for our problem. It may very well be the case that his utility function includes such aims as to earn money, to get promoted, etc., but does not include producing new knowledge. However, it may turn out that he can realise his primary aims only by producing new knowledge and, hence, this becomes his interim aim; *i.e.*, from his utility function and the constraints that obtain in the situation at hand it may follow that producing new knowledge is a suitable, or even an indispensable, means for realising as much as possible of his primary aims. In this case, we get an *indirect* utility function, which includes scientific progress as a means. Since the means will be weighted by the utility function and may change when the constraints change, they are not on the same footing as the primary or "ultimate" aims. The evaluations of the actor are, of course, *subjective*.

The researcher as an ideal type is defined here as a being whose *ultimate aim* is scientific progress; this aim dominates all others. The fact that the property "being a researcher in the sense of the ideal type" may never be exemplified in real life is not a valid objection against the concept. To the extent in which a real person approximates the ideal type, achieving scientific progress constitutes his *intrinsic motivation*. Achieving scientific progress may, usually, be an interim aim. However, since scientific progress can be achieved only through research (a fact which belongs to the ubiquitous constraints), and the direct aim of research is discovery, scientific progress as a direct aim is constitutive of the meaning of research.

The researcher may be viewed as *rational* '*discovery-maximising producer*', who selects and processes problems and problem solutions and, in particular, selects projects on the basis of their expected returns in new knowledge, in discovery. The problem of the choice of discount rate has a counterpart in the individual researcher's activity in the form of his planning and time schedule, his making of a science policy that is *internal* to science. The time preference or "impatience" exactly corresponds to the choice of the discount rate. However, the problem does not arise when we consider basic science as the activity of the idealised researcher, because basic research is an open-ended process. One of the problems of the use of CBA in economics is the problem of which benefits and costs are to be included in the analysis. In the application of CBA-thinking in methodology this problem has a counterpart which leads to the problem of rational theory preference. Also the problem of the valuation of benefits and costs in economic life has a counterpart in the application of CBA in methodology, which likewise leads to the problem of theory preference. In economics any market rate of interest also contains a risk premium. In economic life there sometimes occur situations with completely uncertain outcomes and also situations in which the outcomes themselves may be unknown. In basic research this is the normal situation.

The preference scale of the individual researcher includes, by definition, scientific progress as value or aim; if scientific progress is not among the aims of a particular person, we, *eo ipso*, do not consider him to be a researcher. But the preference scale of any individual will, besides scientific progress, also contain many other values or aims, some of which may not jibe perfectly with the aim of achieving scientific progress pure and simple. Moreover, in the case of an individual, the alternative benefits forgone at the moment of choice are not actually experienced. The decision maker has to imagine them. He compares descriptions or representations of possible situations with expected payoffs. Hence, as already mentioned, the expected effects of the sacrificed alternatives are *subjectively* evaluated (*cf.*, *e.g.*, Pasour [1985]), when the comparative evaluations are made "vicariously", when imagined or expected benefits are being compared in thought experiments.

It is illustrative to distinguish two types of success, *viz.*, success in the sense of an increase in one's reputation and success in the abstract sense, *i.e.*, achieving scientific or intellectual progress, where progress is valued independently of whether or not this achievement will result in an increase in one's reputation. Hence, an action may be recognised as rational, from the viewpoint of a particular researcher, even if it does not realise the scientific progress that the researcher would have been capable of achieving in the long run. Moreover, outlays for having learnt a technique are certainly *past* outlays (and in this sense bygones are forever bygones). However, they are *not necessarily sunk* costs relative to the current decision making of the individual researcher, but may influence his entrepreneurial choice, in particular, if past outlays are high. Thus, it may be rational for him to stick to the old "paradigm", even if he recognises that the new "paradigm" holds more promise in the long run. Retooling to the new techniques may be too time-consuming for him, so that his reputation is best served by continuing to work with the old "paradigm". All this holds good, because, for the individual, the effects of the sacrificed alternatives are *subjectively* evaluated.

When CBA is *applied to the methodology of science as a discipline*, the situation is different. Then, the actor is a construction, an *ideal type*, *viz.*, the researcher *qua* researcher. Personal preference scale and psychological motivation of any particular individual are irrelevant to the methodological issues. Methodology does not aim at describing or explaining the actions of particular scientists or of the scientific community, but it deals with methodological problems. *In research, the basic aim of the activity is given and, hence, the preference structure of the researcher as an ideal type is also given.* The basic aim is scientific or intellectual progress. *This aim is constitutive of the meaning of research* and, hence, it cannot be put into question in that context. Therefore, methodological rules, criteria, decision, etc., are appraised in terms of their *instrumental* value for an approximation to the basic aim, to scientific progress. An outside observer can estimate the subjective evaluations of particular agents, of particular scientists. This falls within the competence of the history of science or the sociology of science. Methodological appraisals are involved, but they are used, *i.e.*, a certain position in methodology is being used as a frame of reference. It is the task of methodology critically to examine the competing methodological appraisals and the criteria that are being used by various historians or sociologists of science. Methodology as a sort of technique for increasing the chances of scientific progress proceeds argumentatively. Like scientific research it stands under the regulative principle of approximating truth and of solving problems that are "scientifically important" in the objective sense. The methodological appraisals associated with the ideal type of researcher or, speaking less metaphorically, with methodology as discipline, as a sort of technology, are objective evaluations. The changing body of scientific knowledge evolves like a spontaneous order (in F. v. Hayek's sense), and through the continuous whittling away of problem solutions that are less good than their competitors, it can achieve progress. It is claimed that CBA applied to the methodology of research may be based on evaluations of costs that are *objective*, in spite of the fact that CBA in personal decision making is bound to the fact that cost as it influences choice is subjective.

1.2 It is high time to clarify my use of '*methodology*'. Methodology is here conceived as a sort of technology for improving the researcher's chances to achieve his aim *qua* researcher: scientific progress. Hence, it constitutes itself an epistemic resource. Methodology is embodied in a set of hypothetical imperatives or recommendations. A typical rule of general methodology may have the following form: 'Presupposing that your aim is scientific progress, new discovery, etc., then, if you find yourself in a problem situation of type S, follow rule R, because by doing so your chances of success will be better than what they would be if you followed any of the rules which compete with R.' Thus, the rules of methodology are basically forward-looking, technological, *i.e.*, hypothetical, prescriptions, rules of action. They include as a subset rules for the *ex post* appraisal of results, of competing problem solutions. A rule of appraisal may have the following form: 'If, at the moment, your problem is to choose between two competing theories, then rule R' provides a rough guidance.' Of course, methodology has also to provide rules for the appraisal of competing methodological rules.

Methodology deals with abstract entities (with Popper's W-3 entities); hence, it proceeds argumentatively. It does *not* attempt to describe or to explain the behaviour of scientists. It is neither an empirical discipline nor a normative enterprise; it is very similar to technology. (*Cf.* Albert [1982], pp. 26–31. This idea is elaborated, *e.g.*, in Radnitzky [1981]—a paper devoted to Popper's methodology—esp. pp. 45–50, in Radnitzky [1982a] and Radnitzky [1985].)

Methodological rules are embedded in a set of global assumptions and may, in turn, influence these assumptions: assumptions about the general characteristics of the human cognitive apparatus, of man's capacity to know (*e.g.*, fallibilism), as well as assumptions about the global characteristics of the aspects of the world that one wishes to investigate: beginning with the assumption of order, the assumption that structure and processes in nature are independent of whether or not we know about them (epistemological realism), the assumption that at least some of the regularities obtained in the the world are discoverable by man, and so forth. Methodology is also interacting with relevant results of empirical inquiry. For instance, for a discussion of basic statements certain findings of psychology are relevant; certain results of linguistics, of mathematics, etc., may help us to improve methodological rules. This interplay of methodology on the one hand, and assumptions of philosophical cosmology and anthropology combined with results of empirical science on the other, does not lead to a vicious circle or to an infinite regress, *if* the efforts to improve methodology are placed in a non-justificationist context, if the quest for knowledge with a truth guarantee has been abandoned (Bartley [1982], [1987a]).

In the course of cultural evolution the criticist tradition developed, presumably with the Pre-Socratics. This tradition emerged as a spontaneous order (as the expression is used by F. v. Hayek), *i.e.*, it was the result of human action but it was not intentionally designed; it arose as an unintended result of purposive actions of individuals who were pursuing their own particular interests. Within the context of the criticist tradition, scientific research, the systematic investigation into nature, emerged as an activity that proceeds by the interplay of creative imagination and systematic criticism, in combination with the provisional retention of those results that have not yet been whittled away by the critical examination. Research is an open-ended process; theories cannot be proved, but sometimes they can be improved. This holds for all sorts of problem solutions. They are fallible, but sometimes we have good reasons for (fallibly) preferring one problem solution to all others, at the moment. The changing body of scientific knowledge as a whole is likewise a spontaneous order; it is the result of a multitude of successful investigations, but it has not been and could not have been designed by a single mind or by any collectivity. (Thus, science as an evolving and changing body of knowledge is an example of a spontaneous order, like language, the market order—Hayek's catallaxy—, morals and law.) Methodology attempts to articulate rules that may increase the researcher's chances to achieve scientific progress. Only if the rules have been articulated, is it possible to examine them critically and, sometimes, to improve them. Methodology attempts, in particular, to articulate and to argumentatively defend rules of rational theory preference and also of rational problem preference. The aim of this essay is to argue that methodology can profit from *explicitly* applying the CBA-frame to its problems.

2 CBA APPLIED TO THE 'PROBLEM OF THE EMPIRICAL BASIS'

2.0 For a Popperian to tentatively accept a statement means that he regards the statement to be worthy to be subjected to further criticism, to severe empirical testing (Popper [1959], New App. *9, p. 419). However, with the so-called basic statements this is different, because they are but raw material in research. With respect to them the relevant questions are: ***When is it rational to regard a basic statement as unproblematic at the moment, and when is it rational to put it into question?***, and, in particular, the question: *How can a basic statement be criticised?*

2.1 Logical positivism and the philosophical tradition in general have been unable to solve the problem of the empirical 'base', because they have placed it in the context of justification philosophy. Popper has solved the problem, at least in principle (Popper [1979 (MS 1930–3)], *Grundprobleme*, and Popper [1934/1959].) Certain aspects of this solution have been criticised, in particular, by W. W. Bartley III. The breakthrough to 'pancritical rationalism' is due to Bartley (*cf.* Bartley [1962], pp. 167 f., Bartley [1984], App. 3, pp. 213 f., and Bartley [1986a]). I hope that by applying CBA to the problem we can improve our grips of it.

According to Popper a basic statement is a report of observation asserting that an observable event is occurring in a certain region of space and time. (The concept is clarified in Popper [1979 (1930–3)], pp. 122, 127, 132: 'Basissätze der Wissenschaft (die "elementaren Erfahrungssätze")'; Popper [1959], Ch. 5, Popper [1963], pp. 386–8; Bartley [1982], pp. 162–8, Bartley [1984], pp. 132–216; Andersson [1984a], pp. 54, 57, 62.) A basic statement must be able to contradict a general statement so that it can function as a potential falsifier of a theory. From the viewpoint of *epistemology*, there is no difference between a basic statement and a general statement: all are fallible in principle, conjectural, criticisable and revisable. From the viewpoint of *methodology*, there is an important difference: basic statements are easier to test than theories and in this sense they are closer to experience. Perceptual experience constitutes the 'cause' and the 'evidencing reason' for our believing and asserting a basic statement.

The tradition that has dominated the philosophy of science has projected a dilemma with respect to basic statements: Either basic statements provide a secure source of knowledge, an epistemic rock-bottom, or they are nothing but the outcome of arbitrary decision or convention. Popper restructured the problem situation by placing the problem into a non-justificationist context (*cf.*, *e.g.*, Andersson [1984a] for a penetrating study of Popper's treatment of basic statements). Popper stresses that observation is used to test. The process of testing is potentially infinite, but since no attempts are made to prove a theory or a basic statement to be false (let alone to prove any of them to be true) no infinite regress need arise. In any field in which experimentation is possible, if we wish to test a theory, we have first to ascertain what sort of events would be incompatible with it, then set up an experimental arrangement with a view to trying to produce such events, and then report the results of the tests (*cf.*, *e.g.*, Popper [1934], Bartley [1982], p. 167, Bartley [1984], p. 215). From the viewpoint of *logic*, a singular statement can falsify a theory. From the viewpoint of *methodology*, a theory is falsified only by a statement that describes a reproducible effect, by a 'falsifying hypothesis'. If a falsifying hypothesis *B* is regarded as less problematic than the theory *T* (plus the statement of the initial conditions and, of course, the relevant auxiliary hypotheses), then *T* is falsified by *B*.

2.2 In the *practice* of research, as well as in daily life, the handling of basic statements does not constitute any problem (*cf.*, *e.g.*, Popper [1959], p. 93, Andersson [1984a], p. 58). From a certain moment we regard a particular statement as unproblematic simply because we are convinced. We cannot decide to be convinced. Experienced certainty is, of course, epistemologically irrelevant, because each and everybody may be wrong. (Popper emphasises this in *Grundprobleme* [1979 (1930–3)], p. 126: 'The important thing is that these *subjective convictions* do not enter into the *scientific procedures of testing*...' [italics in the original, transl. GR].) If a basic statement is not unproblematic at the moment, then a cost-benefit analysis is pertinent, which however, in the practice of research, may remain implicit. How much effort is invested in re-checking a particular statement depends upon the agent's subjective valuation of how much is at stake; *i.e.*, it depends upon his perception of the situation whether he re-checks a statement or turns his attention to other areas. However, to describe the practice of research is not the task of methodology. Therefore, let us turn to the methodological problems.

2.3 While in the practice of research the handling of basic statements does not involve any conscious decision, in the *methodological reconstruction* it is recommendable to reconstruct that handling as a *decision* process. Being unproblematic at the moment means that, at the moment, nothing is known that speaks against it.[1] However, a statement can be unproblematic only in the light of certain other statements that are presupposed, and these are presupposed, because on previous occasions it was decided that they are unproblematic and so far no concrete reasons have emerged for revising these decisions. Since we do not search for an Archimedian point, this need not lead to an infinite regress. Our background knowledge is not normally questioned, and some global assumptions are indispensable as economising devices in dealing with a complex reality.

The decision process involved may suitably be reconstructed as a cost–benefit analysis in which account is taken of the estimated marginal utility of re-checking the statement in question and of the costs involved, in particular, the opportunity costs of the re-checking. From the methodological point of view the important questions are: *How can a particular basic statement be criticised?* and *What are the costs of defending a statement that is 'problematic'?*

---

[1] Popper calls a basic statement 'unproblematic' if there is no reason to test it ('... Basissätze, die nicht weiter überprüft zu werden brauchen.')—Popper [1979/1931–3], p. 132. W. W. Bartley, III, places the distinction problematic/unproblematic in the centre of his study of basic statements. *Cf.* Bartley [1982], sect. XIV and Bartley [1984], pp. 213–16. He dispenses with the concept of decision with respect to basic statements. However, I doubt—for the reasons stated in the text—that in the *methodological reconstruction* we can do without the notion of decision. Perhaps even statistical decision theory may be relevant.

Basic statements are indispensable in testing a theory, but they function only as raw material, *i.e.*, they are not incorporated into the evolving body of scientific knowledge, but serve as a point of departure for producing a falsifying statement. In order to maximise criticism, and thereby the chances of error elimination, it is required that basic statements are inter-subjectively testable. Such testability is a matter of degree. Consider, for example, the report of a pointer reading such as the statement 'At t on this dial the pointer is opposite figure 2' ('b_0' for short). A certain perceptual experience is a motive for asserting b_0 (psychological description), and if the statement is put into question, a report of the perceptual experience—preferably by several observers—will be used to defend it. How much effort will be put into checking the statement depends upon the expected marginal benefits of further research into its truth value. A statement of the type b_0 is unproblematic *only* in the light of certain assumptions such as the assumption that the observer's visual faculty is normal, that parallaxis and glare are under control, etc. If these statements about the initial conditions are not rejected, then to reject b_0 would be tantamount to questioning the general theory of perception that, normally, would be used to analyse perceptual phenomena. This would constitute costs in the objective sense, costs that, normally, are prohibitive.

Statements of the level b_0 are but raw material, *i.e.*, they are of interest only in a problem situation in which from b_0, in combination with statements about the initial conditions of the experimental set-up and universal statements about the functioning of the measuring apparatus used, further test statements are derived such as, *e.g.*, the statement 'At t the pressure in atmospheres in this container is 2' ('b_1' for short). The *predictive implications* of a b_1-statement depend upon the state of the art. Rejecting b_1 while retaining the report about pointer readings b_0 would force us to reject one or more of the other premisses of the argument. If the basic statement about the pressure is rejected while the statement about the pointer reading is retained, and if it is also claimed that the instrument used functioned properly, then one would be forced to reject the physical theory underlying the technology in accordance with which the measuring instrument has been constructed. If there are no concrete reasons for problematising that theory, this would amount to an irrational allocation of scarce resources of time and effort.

What does it mean to re-check, to test, a basic statement like b_1? It will mean measuring again, perhaps with another measuring technique, getting others to check the functioning of the various measuring apparatuses, and so forth. Whether or not investing in re-checking a particular basic statement is a good investment decision depends in large measure on the objective problem situation. If a basic statement *b* corroborates the theory *T* that we wish to test (*i.e.*, if the situation is that of an unsuccessful falsification attempt), then the cost of repeating the experiment as well as the expected marginal utility of doing so are likely to be small or negligible. If *b* contradicts *T* (given the statements of the initial conditions and the relevant auxiliary hypotheses), then investing time and effort into re-checking *b* holds more promise of an intellectual gain. Thus, the problem situation is decisive. However, in general, re-testing a b_1-statement is a profitable enterprise only when we wish to produce a falsifying hypothesis, *i.e.*, a statement of a form: 'Whenever certain initial conditions arefulfilled, then b_1 (the pressure in the container will be 2 atmospheres)' ('*B*' for short). A statement of the *B*-type is of interest if it contradicts the theory that we wish to test. Only a *B*-type statement possesses sufficient intersubjective testability to function as a falsifying hypothesis.[1] Therefore, the methodological problem of the empirical 'base' turns out to be essentially the problem of whether or not to ***invest*** time and effort into ***processing a particular basic statement into a falsifying hypothesis for the theory we wish to test***. Like any other investment decision, this decision can rationally be made only on the basis of a CBA. The decisive factor is the objective problem situation. Thus, in the above example of a b_0-statement, one of the options that, in principle, are open to the researcher is to question one of the global assumptions of the background knowledge (*e.g.*, the theory of perception involved in the argument), instead of re-checking the b_0-statement or questioning one or more of the statements of the initial conditions or the theories involved. The cost of such a move is likely to be very high, the chance of success small—whether there is any chance at all depends on the particular problem situation—, while the payoff, if one should succeed, may be great. Thus, it is rational to embark on such an enterprise ***only if*** there are ***concrete*** reasons for challenging that particular global assumption of the background knowledge.

To sum up, in the practice of research the problem of the empirical 'base' is no problem. In the methodological reconstruction of the handling of basic statements it is advisable to reconstruct that handling as involving essentially an investment decision based on CBA. The problem of the empirical 'base' does not constitute any major problem of methodology. Rather it is an epistemological problem that cannot be avoided, because we test theories by means of test statements and test statements themselves by deducing from them other basic statements. Knowledge in the objective sense is possible only because there are basic statements, test statements, that are regarded as unproblematic, at the moment (Popper [1979 (1930–3)], p. 132).

---

[1] Popper [1979/1930–3], p. 132: '... and this is the decisive point—that what matters in it (in scientific method) ... is primarily not particular, ***singular statements*** (***besondere Sätze***), but ***general regularities*** (***allgemeine Gesetzmäßigkeiten***); *i.e.*, statements that ... can and must be tested again and again in a deductive manner (by deducing many and very different singular statements)' (italics in the original, Engl. transl. GR). See also Andersson [1984a], p. 63.

3 CBA APPLIED TO THE PROBLEM OF RATIONAL THEORY PREFERENCE

3.0 In technological *applications* of scientific theories, and in everyday life, the pertinent question is indeed whether or not to accept a theory. We accept a theory in the sense that we trust it will help us to realise a particular practical goal. This expectancy is rational only if the theory is sufficiently well-tested for the practical purpose at hand. For economic reasons it is often rational not to use the best (in the methodological sense) of the available theories. (On the relationship between science and technology see, *e.g.*, Radnitzky [1983], pp. 235–43.)

In *research* the situation is completely different. The question of when it is rational to stop problematising a particular position is a reasonable question if asked with respect to a basic statement, because basic statements are but raw material of science. If the question is asked with respect to theories, it is misleading because it is based on the false assumption that research has a 'natural' end. Basic research is open-ended, because solved problems create new problems (Popper [1959], p. 104; Popper [1934], pp. 69 f.). In the context of basic research to speak of 'accepting' a theory is innocuous only if by 'accepting' we mean that we regard the theory as worthy of being subjected to further empirical testing (Popper [1959], New App. *9, p. 419). In basic science predictions are but a means of testing a theory. (On the distinction between basic and applied research *cf.*, *e.g.*, Radnitzky [1983].) Exempted from criticism, *pro tempore*, are only those theories that we use as auxiliary hypotheses when testing a theory or when producing an explanation. In sum, so long as a theory is taken seriously as a possible solution to a scientifically interesting problem, it has to be tested.

When a theory has been falsified, the researcher has the choice between investing time and effort into modifying the 'old' theory or investing his time in replacing it by a radically different one. In each case there is competition of theories: either between the 'old' theory and the improved successor (which, although it is but a modified version of the 'old' theory, is intended to replace it) or between the 'old' theory and the radically new theory. Hence, the issue is not one of acceptance or rejection of a single theory, but rational theory *preference*. We always start with some background knowledge, some assumptions about the sort of phenomena to be explained, *i.e.*, some theory or protoform of a theory; in this sense there is some competition of theories in every problem situation. The question of whether a particular theory is preferable to another arises only if the two theories are competitors. Theories are competitors (in a narrow sense of the word) if they offer incompatible solutions to the same, or highly overlapping, sets of explanatory problems; they are competitors in the wide sense if their realms of applicability overlap partially, but one theory has greater explanatory power than the other.

Popper restructured the problem situation by placing the problem of theory appraisal in the non-justificationist context. The style of thought that has dominated the philosophy establishment, logical empiricism, has projected an ideal of science according to which progress consists in probabilifying a theory more and more. Those who recognised that the logical positivists' program is a will-o'-the-wisp but could not free themselves from the justificationist ideal of knowledge, either adopted the instrumentalist view of theories or joined the sociological turn in the 'philosophy of science'. (For a critique of this development see, *e.g.*, Part III of Radnitzky and Bartley [1987].) The sociological turn is at the same time a relativistic turn in the so-called sociology of knowledge: what matters is success in the sociological sense; there is no other criterion, because theories belonging to one 'paradigm' are incommensurable with those developed under another 'paradigm'. Thus, theory instrumentalists and relativists have lost or abandoned the problem of rational theory preference. Popper, on the other hand, replaced the question 'How can we prove a theory?' by the question 'How can we improve a theory?'

3.1 *From the Problem of Rational Theory Preference to the Problem of a Research Policy that is Internal to Science*

What is the rational response to a falsification? Is the CBA-frame relevant in that problem situation? A successful falsification attempt of a theory T results in the insight that from certain premises, which we regard as unproblematic at the moment, it follows that non-T. This is all *logic* can tell. Given this insight and the basic aim of research, the rational response to the problem situation is governed by the following global methodological rule: Propose a theory that is not hit by the falsifying hypothesis in question; you may eliminate the part that you blame for the falsification—but, if you did so repeatedly, the theory would be emptied of its empirical content (and the costs of such immunising tactics would be too high). This is all *methodology* can tell.

Now the researcher is faced with a problem of a research policy that is internal to science: to allocate scarce resources of time and effort within a given field of study.[1] For this sort of *risky investment decision* it is in principle impossible to give any methodological rule, if only because it is impossible to know in advance whether a modified version of the falsified theory will do, whether we will be able to create a radically new problem solution, and so forth. In this problem situation the CBA-frame can help the researcher to see which sorts of questions he should take into account when comparing competing investment projects, what good reasons can be given for a particular problem preference, and what good reasons can be given for preferring a particular theory to its competitors.

---

[1] Science policy that is *internal* to science must be clearly distinguished from science policy that is *external* to science: the political decisions allocating financial and other resources among fields of study, of setting budgetary frames, etc. For an overview *cf.*, *e.g.*, Radnitzky [1983].

What good reasons can be given for the methodological decision or verdict that a theory T' is to be *preferred* to a theory T? Time and again the researcher has to answer this question. He has to function as a methodologist of his own. (Sometimes, he may not be aware of this fact (*cf.*, *e.g.*, Radnitzky [1981], pp. 45–50). Incidentally, those who deny that they operate within some methodological framework and claim that they do not need any methodology are as a rule adherents of naive positivism.) In order to clarify what should be meant by 'good reasons' in this context one must, first, explicate the idea that a theory T' is better than T in certain respects, *i.e.*, make explicit the relevant dimensions of criticism;[1] second, produce arguments for the claim that it is these criteria that matter, *i.e.*, that they follow from the basic aim of science; third, state what empirical support can be given for the conjecture that T' fulfills these requirements in a higher degree than T does.

Comparison of past performance in explaining and predicting enables us to make a conjecture about the problem-solving capabilities of the competitors. That means that we have to *compare the two balance-sheets of past performance*. Thereby all available information should be used, including appraisals of the 'scientific importance' of the problems the theory can solve or has failed to solve. Following Popper's methodology we conjecture that the theory that has so far survived all the tests or more of the tests than the rival theory possesses a greater problem-solving capability than the rival that performed less well. On a realist view of theories we attribute this to the fact that the more successful theory has a greater capability of describing certain aspects of reality, that it is more truthlike than its rival. Since the logical problem of induction cannot get a positive solution, conjectures about future performance of theories on the basis of their differential past performance will always be very risky.

Since in the methodological appraisal of the past performance of theories not only the logical point of view is relevant, it would not be rational to consider equivalent a failure to explain a fact that is scientifically fairly unimportant and a failure to explain an observed regularity of great scientific importance. Hence, success and failures should be weighted by the scientific importance of the problems concerned. The concept of 'scientific importance' can be explicated objectively in terms of the contribution which a correct solution to a problem makes for progress in the discipline concerned. (*Cf.*, *e.g.*, Radnitzky and Andersson [1979], pp. 244–9; Radnitzky [1980], pp. 221–7.) This explicatum has nothing to do with the subjective interests or predilections of individual scientists. In sum, through the appraisal *ex post* of the relative scientific importance of the problems that one of the competing theories solved while its rival did not, the problem of rational theory preference includes part of the problem of *rational problem preference*. The appraisal, *ex ante*, of the relative scientific importance of competing problems is a problem that confronts every scientist who works on his own responsibility, and it is also the key problem of a research policy that is internal to science.

---

[1] On the dimensions of criticism of theories, *cf.*, *e.g.*, Radnitzky [1974], esp. fig. 4 on p. 90; Bartley [1982], pp. 161–77. The various dimensions of criticism have to be stated and ranked in a way that fits the particular problem situation. Space does not permit me to elaborate this. Let me just mention that the basic pre-testing criterion is relative content of empirical information—how much a theory says. Popper recommends bold conjectures, *i.e.*, he recommends an *entrepreneurial spirit* in research. *Like in business, in science risk and possible gains are correlated.* (The basic post-testing criterion is, of course, whether what the theory says, or how much of what it says, is accurate.)

On the micro-level, the individual researcher acting as a discovery-maximising entrepreneur, who orients himself on the estimated returns in new knowledge, has to face the problem of rationally allocating scarce resources of time and effort among the problems that compete for his attention. These allocations are risky investment decisions. To appraise whether or not such an investment decision has been well-taken, to monitor this allocation activity, one needs the methodological rules of theory appraisal, in particular, of theory preference. On the macro-level, in basic science, the government agency that has been entrusted with science policy has to set budgetary frames for the various disciplines, and then the subgroup of the scientific community that represents the discipline in question has to allocate the scarce resources of talent, time and effort among the problems that compete for attention, at the moment, *i.e.* in view of the problem situation in the discipline. In making these risky investment decisions, both the researcher and the peer group may be helped by the CBA-frame. The CBA-frame will draw their attention to such questions as the question of the costs of defending a theory whose past performance is less good than that of its competitor, to the problem of the costs of acquiring certain epistemic resources such as methods, skills, conceptual frames, and so forth. The rational option for one of two competing theories, or for one of two competing problems, has to be based upon a CBA where the valuation of benefits and costs is *objective*. In all those fields of studies in which theorising is closely monitored by empirical testing, theory change is an objective process. In such fields the better theory will eventually drive out the less good one, and do so for economic reasons. (An analogy from economics could be that in a free market for money high confidence money would drive out low confidence money.)

3.2 *CBA Can Show that Theory Change is an Objective Process, in which the Evaluation of Benefits and Costs is Objective.*

The best way of arguing for this thesis is to show CBA in operation. I adapt the simple example from the history of science that George Schlesinger uses in his review of Thomas Kuhn's *The Essential Tension* (Schlesinger

[1981], pp. 458 f.). Schlesinger uses the example of the round-earth theory superseding the flat-earth theory to show that theory change is an objective process. Since ancient times there were observations (basic statements) that provided the raw material for a falsifying hypothesis of the flat-earth theory, above all the repeatable observation that the sails of an approaching ship can be seen before the ship's hull can be seen. This assertion is not only sufficiently neutral with respect to the two competing theories to serve as a crucial experiment, but the observation can be made by anybody independent of his world-picture, the nature of his language, etc. With the help of the round-earth theory the observation is readily explained. The flat-earth theorist can immunise his theory against the falsification, but only at a certain cost, *e.g.*, by introducing certain *ad hoc* hypotheses. (On *ad hoc*-ness *cf.*, *e.g.*, Radnitzky and Andersson [1979], pp. 242–4.) With the help of the round-earth theory it can readily be explained why lunar eclipses are round. The flat-earth theory can be defended only at high costs, *e.g.*, by denying that lunar eclipses are due to the shadow that the earth casts on the moon. This immunising move is particularly costly, because as a consequence of it one is obliged to reject also the theory that explains, *e.g.*, why lunar eclipses occur in the middle of the lunar month. The first sailing around the earth can readily be explained by the round-earth theory. The costs of defending the flat-earth theory become increasingly high: the fact has to be explained away, one has to claim that in spite of all appearances no circumnavigation did take place. Eventually, the costs of defending the falsified theory become unbearably high, *e.g.*, when confronted with pictures of the earth taken from satellites. In all these cases the relevant valuation of the costs is objective. It is easy to cull from the history of science examples that illustrate this process: the Ptolemaic theory versus the Copernican, the hollow-earth theory versus the Copernican theory (*cf.*, *e.g.*, Sexl [1983]), the theory of the ether (at least in the form given it in a Newtonian framework) versus the theory of relativity, and so forth.

Let me just mention an often-used example that is more modern. Since the middle of the nineteenth century it has been known that the perihelion of Mercury deviates from what is calculated with the help of Newton's theory. Newton's theory can be defended against this falsification, but only at a certain cost, *i.e.*, by introducing certain *ad hoc* modifications of the auxiliary hypotheses: either by assuming an additional planet or by assuming that the mass in the sun is distributed unevenly. Each of these *ad hoc* assumptions is testable, but neither could be corroborated. In 1916, an accurate description of the orbit of Mercury was deduced from the statement about the initial conditions of the planetary system (without *ad hoc* assumptions) with the help of the general theory of relativity. From this moment the costs of defending Newton's theory increased dramatically. Beside the costs of the above mentioned *ad hoc* modifications (which remained *ad hoc* since they could not be corroborated) there is the additional cost of having to reject not only the accurate derivation of

Mercury's orbit that can be produced with the help of the general theory of relativity but, *by implication, also to reject the theory that changed our world view*. Hence, in this case the objective costs of defending the less good theory suddenly became unbearably high.

The fact that during an initial period it may not be possible to muster good reasons for preferring one of the competing theories (something which Paul Feyerabend stresses) is not a valid objection against the thesis that the valuation of costs of defending the less good theory are objective and, eventually, will become prohibitive.

3.3 *A Remark on the Costs Associated with* ad hoc *Hypotheses*

When a theory has been falsified, the researcher has the choice of modifying the theory or—if he is sufficiently creative and lucky—replacing it by a radically different one (*cf.* section 3.0 above). He may modify the theory *ad hoc* or make *ad hoc* modifications in the premises of an explanatory argument (as in the Mercury example). What are the costs and possible benefits in such cases? *Ad hoc*-ness is a notoriously difficult concept. In many cases the resource cost of 'repairing' the situation (falsification situation) by means of an *ad hoc* modification of the premises is negligible but the payoff is at best spurious. What matters is (*pace* Popper) whether or not the *ad hoc* modification is testable *as a part* of the theoretical system constituted by the theory to be tested, the initial conditions and the auxiliary hypotheses (cf. Andersson [1987], chapter 7). If the hypothesis that was *ad hoc*-ly introduced should be corroborated, then what originally was an 'as-if'-explanation would become an authentic explanation. At this juncture the 'time' factor becomes important. If one has made an *ad hoc* addition, one is obliged to make efforts to test that modification. If one is slow to make the necessary investment in time and effort, then the opportunity costs are bound rapidly to increase. If the *ad hoc* modification is retained in spite of the fact that it has been falsified, then the opportunity costs become forbidding, because proceeding in this manner would stop scientific and intellectual progress.

4 CONCLUSIONS

What matters in the first place is whether or not a statement—be it a 'basic' statement or a theory—is true. In the case of 'basic' statements the key problem is a practical problem: efficiently to allocate scarce resources of time and effort. The costs of a less good investment decision are opportunity costs: possible gains in new knowledge forgone, *i.e.*, the potential loss of knowledge, knowledge that might have been achieved if the resources had been invested in another project than the one chosen. When making an investment decision the CBA-scheme is indispensable. In the case of theories the problem is a methodological one. The cost of preferring the theory that has less explanatory power than its rival is the loss of epistemic resources. Epistemic resources that already exist in the market of ideas suddenly become unavailable. This sort of cost is objective. In rational theory preference the CBA-frame is helpful. This may seem trivial since we certainly make use of CBA whenever we act rationally. Making CBA-considerations *explicit* helps us to frame the problem situation, to articulate the various steps in the decision-making process and thereby to make them more criticisable, and also to see what the options are. Hence, the CBA-frame constitutes itself an epistemic resource for the researcher whenever he has to act as a methodologist of his own. The claims defended here are, firstly, that CBA is not only compatible with Popper's methodology, but that it may pay to view that methodology as an application of CBA to epistemic situations and, secondly, that the program of systematically exploring the applicability of CBA to methodological problems may help us to improve our understanding of the philosophy of science.

University of Trier, West Germany

REFERENCES

ALBERT, H. [1982]: *Die Wissenschaft und die Fehlbarkeit der Vernunft.* Tübingen: Mohr.

ANDERSSON, G. [1984a]: 'How to Accept Fallible Test Statements? Popper's Criticist Solution' in Andersson [1984b], pp. 47–68.

ANDERSSON, G. (*ed.*) [1984b]: *Rationality in Science and Politics.* (*Boston Studies in the Philosophy of Science*, 79). Dordrecht: Reidel.

ANDERSSON, G. [1987] (*in press*): *Kritik und Wissenschaftsgeschichte.* Tübingen: Mohr.

ANTISERI, D. [1981]: *Teoria Unificata del Metodo.* Padova: Liviana.

BARTLEY, W. W., III [1962]: *The Retreat to Commitment.* New York: Alfred A. Knopf, Inc., and London: Chatto and Windus Ltd. Second Edition, revised and enlarged [1984]. La Salle, IL: Open Court.

BARTLEY, W. W., III [1982]: 'The Philosophy of Karl Popper: Part III: Rationality, Criticism, and Logic', *Philosophia (Israel)*, **11**, pp. 121–83.

BARTLEY, W. W., III [1987a]: 'Philosophy of Biology *versus* Philosophy of Physics', in Radnitzky and Bartley [1987], pp. 7–45.

BARTLEY, W. W., III [1987b]: 'Alienation Alienated: The Economics of Knowledge *versus* the Psychology and Sociology of Knowledge' in Radnitzky and Bartley [1987], pp. 423–51.

BARTLEY and RADNITZKY—see Radnitzky and Bartley [1987].

BROWN, C. and JACKSON, P. [1978]: *Public Sector Economics.* Oxford: Martin Robertson.

GHISELIN, M. [1974]: *The Economy of Nature and the Evolution of Sex.* Berkeley: University of California Press.

GHISELIN, M. [1978]: 'The Economy of the Body', *The American Economic Review* **68**, pp. 233–7.

GHISELIN, M. [1987]: 'Principles and Prospects for General Economy', in Radnitzky and Bernholz [1987].

GRAY, J. [1987]: 'The Economic Approach to Human Behavior: Its Prospects and Limitations', in Radnitzky and Bernholz [1987].

GRMEK, J., COHEN, R. and CIMINO, G. (*eds.*) [1981]: *On Scientific Discovery. The Erice Lectures 1977* (*Boston Studies in the Philosophy of Science*, **34**). Dordrecht: Reidel.

GROVE, J. [1982]: 'The Sociological Denigration of the Rationality of Science', *Minerva*, **20**, pp. 550–6.

HANSMEYER, K.-H. (*ed.*) [1971]: *Das rationale Budget. Ansätze moderner Haushaltstheorie.* Köln: Kölner Universitätsverlag.

HAYEK, F. v. [1973, 1976, 1979]: *Law, Legislation and Liberty.* 3 Vols. London: Routledge and Kegan Paul.

HEINRICH, B. [1979]: *Bumblebee Economics.* Cambridge, MA.: Harvard University Press.

HIRSHLEIFER, J. [1977]: 'Economics from a Biological Viewpoint', *The Journal of Law and Economics*, **20**, pp. 1–52.

HIRSHLEIFER, J. [1980]: *Price Theory and Application.* Englewood Cliffs, NJ: Prentice Hall.

HIRSHLEIFER, J. [1982]: 'Evolutionary Models in Economics and Law', *Research in Law and Economics*, **4**, pp. 1–60.

HIRSHLEIFER, J. [1985]: 'The Expanding Domain of Economics', *The American Economic Review*, **75**, pp. 53–68.

HIRSHLEIFER, J. [1987]: 'The Economic Approach to Conflict', in Radnitzky and Bernholz [1987], pp. 335–64.

KIRSCH, G. [1971]: 'Die Cost-Benefit-Analyse: Ein Katalog von Fragen', in Hansmeyer [1971], pp. 55–104.

NEUMAIER, O. (*ed.*) [1986] *Wissen und Gewissen.* Wien: VWGÖ Verlag.

PÄHLER, K. [1986]: *Qualitätsmerkmale wissenschaftlicher Theorien.* Tübingen: Mohr.

PASOUR, E. C. [1985]: 'Opportunity Cost, Sunk Cost, and Entrepreneurial Choice', *Economia delle Scelte Pubbliche* **3**, pp. 19–30.

PETRONI, A. M. [1981]: *Karl R. Popper: Il Pensiero Politico.* Firenze: Le Monnier. With Introduction by Petroni, pp. 5–110.

POPPER, K. [1957]: *The Poverty of Historicism.* Tenth Impression [1979]. London: Routledge and Kegan Paul. Originally published in *Economica* [1944/45].

POPPER, K. [1959]: *The Logic of Scientific Discovery.* London: Hutchinson. 11th impression [1982]; Originally published as *Logik der Forschung.*. Wien: Springer [1934].

POPPER, K. [1967]: 'La Rationalité et le Statut du Principe de Rationalité', in Classen, E. (*ed.*), *Les Fondements Philosophiques des Systèmes Economiques.* Paris: Payot, pp. 142–150. Shortened Engl. transl. in Miller, D. (*ed.*) [1983]: *A Pocket Popper*, London: Fontana, pp. 335–65.

POPPER, K. [1972]: *Objective Knowledge.* Fifth Edition, revised [1979]. Oxford: Clarendon Press.

POPPER, K. [1976]: *Unended Quest. An Intellectual Autobiography.* London: Fontana.

POPPER, K. [1979]: *Die beiden Grundprobleme der Erkenntnistheorie.* Tübingen: Mohr.

POPPER, K. [1984]: 'Against Induction: One of Many Arguments', in Andersson [1984b], pp. 245–8.

RADNITZKY, G. [1964]: 'The Internal Rate of Profit as a Measure of Profitability', *Erhvers-økonomisk Tidskrift* (Copenhagen), **28**, pp. 195–211.

RADNITZKY, G. [1974]: 'From Logic of Science to Theory of Research', *Communication and Cognition*, **7**, pp. 61–124.

RADNITZKY, G. [1980]: 'From Justifying a Theory to Comparing Theories and Selecting Questions', *Revue Internationale de Philosophie*, **34**, pp. 179–228.

RADNITZKY, G. [1981]: 'Progress and Rationality in Research', in Grmek *et al.* [1981], pp. 43–102.

RADNITZKY, G. [1982]: 'Teoria della Scienza', *Enciclopedia del Novecento*, Roma: Istituto della Enciclopedia Italiana, Vol. VI, pp. 370–86.

RADNITZKY, G. [1983]: 'Science, Technology, and Political Responsibility', *Minerva*, **21**, pp. 234–64.

RADNITZKY, G. [1985]: 'Réflexions sur Popper.—Le Savoir, Conjectural mais Objectiv, et Indépendent de toute Question: Qui y Croit? Qui est à son Origine?' *Archives de Philosophie* (Paris), **48**, pp. 79–108.

RADNITZKY, G. [1986]: 'Responsibility in Science and in the Decisions about the Use or Non-Use of Technologies', in Neumaier [1986], pp. 99–124.

RADNITZKY, G. [1987a]: 'In Defense of Self-Applicable Critical Rationalism', in Radnitzky and Bartley [1987], pp. 279–312.

RADNITZKY, G. [1987] (*forthcoming*): 'An Economic Theory of the Rise of Civilization and its Policy Implications: Hayek Account Generalized', *Ordo*, **38**.

EADNITZKY, G. and ANDERSSON, G. (*eds.*) [1979]: *The Structure and Development of Science* (*Boston Studies in the Philosophy of Science*, Vol. **59**). Dordrecht: Reidel.

RADNITZKY, G. and BARTLEY, W. W., III (*eds.*) [1987]: *Evolutionary Epistemology, Rationality and the Sociology of Knowledge*. La Salle, IL: Open Court.

RADNITZKY, G. and BERNHOLZ, P. (*eds.*) [1987]: *Economic Imperialism. The Economic Approach Applied outside the field of Economics*. New York: Paragon House Publishers.

SCHILPP, P. (*ed.*) [1974]: *The Philosophy of Karl R. Popper (The Library of Living Philosophers)*, 2 Vols. La Salle, IL: Open Court.

SCHLESINGER, G. [1981]: 'The Essential Tension: Selected Studies in Scientific Tradition and Change, by Thomas Kuhn. The University of Chicago Press, 1977', *Philosophia* (Israel), **9**, pp. 455–67.

SEXL, R. [1983]: 'Die Hohlwelttheorie', *Der Mathematische und Naturwissenschaftliche Unterricht*, **36**, pp. 453–60 (Heft 8).

Brit. J. Phil. Sci. 38 (1987), 181–205 *Printed in Great Britain*

Lakatosian Heuristics and Epistemic Support

by THOMAS NICKLES*

1 *Introduction*
2 *Lakatos on Heuristic vs. Justificationist Methodology*
3 *Why Is Lakatosian Heuristics Important?*
4 *Generativism and Consequentialism*
5 *The 'Parameter Fixing' Argument*
6 *The Incongruity of Lakatosian* Adhoc-*ness Rules*
7 *Is Lakatosian Methodology Incoherent?*
8 *Empirical Support, Generative and Consequential*
9 *Heuristics and Generative Support*

1 INTRODUCTION

Although Karl Popper frequently remarked on the heuristic importance of metaphysical ideas in science, he did not develop a full-fledged, heuristic methodology of science. Problem solving heuristics[1] fall between the two stools of Popper's permissive 'method of conjectures and refutations' and his strict view that methods are general algorithms (and hence that, properly speaking, there is no scientific method). It was Imre Lakatos, among major philosophers of science and mathematics, who made heuristics respectable. Latter day Lakatosians such as Elie Zahar, John Worrall, and Peter Urbach have further developed Lakatos's heuristic 'methodology of scientific research programmes' (MSRP). Zahar [1983] in particular has attempted to rescue Lakatosian heuristics from a limbo between logic and psychology by showing that heurisics can be rational and even deductive.

Despite these advances on Popper's methodology, I contend that Lakatos and his successors have not gone far enough in the direction of a heuristic, generative methodology. Because MSRP retains a Popperian conception of empirical support, it fails to provide an adequate motivation for heuristics and is ultimately incoherent.

---

* I am grateful to Andrew Lugg and James Woodward for criticism of an early draft and to the U.S. National Science Foundation for research support.

[1] When I employ singular constructions such as 'heuristics is . . .', I mean heuristics as a field of study, a subject matter. When I use the plural ('heuristics are . . .'), I mean individual heuristic procedures.

2 LAKATOS ON HEURISTICS *vs.* JUSTIFICATIONIST METHODOLOGY

As Popperians and Lakatosians use the term, a 'justificationist' theory of knowledge is one committed to the existence of foundations of knowledge, at least probabilistic foundations. Justificationism includes inductivism and all forms of *generativism*—the view that certain modes of generating (constructing) a claim provide empirical support for that claim. Let us call any such support *generative support*. As I shall employ the term, a *constructive* methodology is one which attaches importance to the manner in which hypotheses and other research products are constructed, but *without* necessarily attaching any *epistemic* significance to the mode of construction. Lakatos was a constructivist but not a generativist. Popper is neither.

In a revealing passage, Lakatos wrote:

> [I]t was only Popper who recognized that the *prima facie* inconsistency between the few odd, casual remarks against *ad hoc* hypotheses on the one hand and the huge edifice of justificationist philosophy of knowledge must be solved by demolishing justificationism and by introducing new, nonjustificationist criteria for appraising scientific theories based on anti-adhocness ([1970], p. 124).

Lakatos contended that earlier methodological thinking confusedly attempted to graft together mutually incompatible methodologies of science. Ironically, Lakatos himself, in following Popper here, made the same mistake. For MSRP attempts to graft together two incompatible methodological traditions. On the one hand, Lakatos retained the idea of the generative tradition that new theories should be motivated by, or even derived from, a body of previous results and plausible guiding principles, which Lakatos termed the 'hard core' and the 'positive heuristic' of the programme. On the other hand, he embraced Popper's view that successful novel prediction is the *only* source of empirical support in science; hence, the manner in which scientific claims are (or can be) generated never furnishes any epistemic support whatever.

In deploring any merging of heuristics and epistemology, Lakatos apparently feared that his readers would confuse his heuristics with the foundational methodologies of the seventeenth century—and with inductivist epistemologies in particular. But an epistemic heuristics need be neither foundational nor inductivist. Not all epistemic considerations are foundationist; otherwise, Popper's fallibilistic epistemology would be impossible. There is a clear sense in which Popperian corroboration and falsification are epistemic processes and products, in contrast, say, to psychological inspiration and aesthetic preference. We can speak perfectly well of empirical evidence and other forms of epistemic support without commitment to 'positive support' of a kind Popperians reject.

My question, then, is: Why cannot the constructive component of methodology be epistemically relevant in this fallibilistic sense, if the test-consequential component can be? We cannot exclude this possibility *a priori*. To observe that *classical* generative programmes (discovery methodologies) were foundationist does not entail that *all* versions of generativism are foundationist any more than Whewell's infallibilism tainted all future doctrines of novel prediction. If consequential support can be nonfoundational, so can generative support.

It follows that arguments against foundationism and against inductivism are not necessarily arguments against generative methodologies *überhaupt*. Lakatos's grand division of methodologies into epistemological *versus* heuristic methodologies was a false dichotomy.

Popper's methodology shows how we can be empiricists (how we can learn from experience) without being inductivists. But Popper mistakenly concluded that scientists never learn from experience by constructive means and that all constructive methodologies are generative, hence justificationist, and thus untenable. (See, *e.g.*, his instructive [1963b].) Accordingly, Popper's methodology is purely critical and eliminative rather than constructive. Lakatos attempted to go beyond Popper, to show how to be a constructivist without being a generativist. Zahar in turn showed that Lakatos's heuristic constructivism is logically coherent in not backsliding into inductivism or psychologism. But because Lakatos, Zahar *et al.* followed Popper in conflating anti-justificationism in epistemology with the view that the constructive component can have no *epistemic* role in inquiry, they launched MSRP away from Popper's position only to orbit around it. I should like to provide the additional fuel necessary to achieve escape velocity. I attempt to improve on MSRP by showing how to be a generativist without being a justificationist and how to combine generativism and consequentialism without backsliding into inductivism, psychologism, *or* into an historical conception of justification.

3 WHY IS LAKATOSIAN HEURISTICS IMPORTANT?

Why do Lakatos and his successors[1] think heuristics important? What intellectually motivates and what justifies their emphasis on heuristics? At the most general level, Lakatos saw only two possible bases for methodology: foundational epistemology and heuristics. Since foundational epistemology has succumbed to fatal criticism, he reasoned, that leaves the field to heuristics, in the broad sense of 'the rational reconstruction of the growth of scientific knowledge'. This unusually broad use of 'heuristics' is misleading, since heuristics are supposed to be forward-looking and constructive, not retrospective and reconstructive. Ironically, in this broad sense, even Popper's methodology is heuristic. Clearly, the failure of foundational epistemology does not guarantee the methodological importance of heuristics in the strict, literal sense of 'heuristics'.

---

[1] See especially Zahar [1983], Worrall [1978a], and Urbach [1978]. They sometimes disagree with Lakatos and among themselves.

Why then do Lakatos and his successors think heuristics *sensu stricto* important to methodology? Their writings reveal a variety of reasons. (1) A heuristic approach is pedagogically superior—less artificial and authoritarian than the 'Euclidean' approach (Lakatos [1976, 1978b]). (2) A heuristic methodology provides a more accurate, complete, and rational account of the historical growth of knowledge and is thus more historically testable than its competitors (Lakatos [1971]). (3) A heuristic methodology may therefore provide an explanation of scientific progress (*cf.* Zahar [1983], p. 246). (4) Heuristic methods improve problem-solving efficiency, *e.g.*, by transforming the analysis of naive conjectures into research programmes, which, *inter alia*, uncover hidden lemmas. (5) Hence a programme with a strong heuristic is likely to produce new theories and predictions *and* (6) to anticipate and overcome anomalies at a more rapid rate than a weaker competitor. (7) The more heuristic power a programme possesses, relative to its competitors, the better justified the conventional decision to make its hard core unfalsifiable (see Musgrave [1978], pp. 188 ff.). However, the Lakatosians insist, we must never confuse heuristic power with epistemic warrant. Heuristic considerations provide zero epistemic support.

Zahar [1983] argues that heuristics are more than psychological crutches. Heuristics can be rational, and, indeed, deductive. Yet Zahar goes no further than Lakatos in telling us precisely what the rationale of heuristics is. He does say, in so many words, that 'rational' heuristics does *not* mean 'epistemically relevant' heuristics. Why then include heuristics in methodology? Does it fulfil some essential, nonepistemic function?

Epistemic relevance is only one rationale for including something in methodology. Economy of research (cost-benefit, efficiency, and sufficiency or 'satisficing' considerations) provides a second rationale. Since Lakatosians adamantly refuse to assign heuristics a direct, epistemic role in inquiry, the rationale of heuristics for them is presumably economic.

Actually, as (4)–(7) *above* suggest, Lakatosians *can* say that heuristics has epistemic relevance, in an indirect sort of way: possession of an efficient heuristic contributes to the positive appraisal of a research programme by improving its capacity to produce novel results, some of which may happen to be corroborated, and also by holding off epistemic criticism as premature. As long as the programme retains sufficient heuristic power, the day of epistemic reckoning is postponed indefinitely. However, in terms of providing a direct warrant for the claims it produces, Lakatos's answer was clear: heuristics, including all rules and information employed in theory construction, carries *zero* epistemic weight. The process of production does not at all justify the product; moreover, any information used in the production process carries zero weight even though it might have possessed nonzero weight had it been employed *post hoc*, in predictive testing.[1]

Although Lakatosian heuristsics has an economic rationale, I shall argue that it needs a directly epistemic rationale also, by the Lakatosians' own principles. However, we must first gain a clearer understanding of the alternatives.

4 GENERATIVISM AND CONSEQUENTIALISM

The Lakatosians appear to be generativists until it comes to epistemic support, whence they abruptly revert to the pure consequentialism usually associated with hypothetico-deductive (H-D) methodologies. To proceed further, we must distinguish several varieties of generativism and consequentialism. Failure to draw these distinctions has spoiled most treatments of this subject.

Generative arguments reason *to* the claim in question as a conclusion from warranted premises; consequential arguments reason *from* it to its testable consequences. Hence, generative arguments explore sufficient conditions of a claim, while consequential arguments probe its necessary conditions. Consequentialism is not a controversial position today. Generativism is.

It is important to see that one may consistently be both a generativist and a consequentialist; however, not all varieties of generativism are compatible with all varieties of consequentialism. As Laudan [1980] observes, most classical exponents of discovery programmes considered generative support more important than consequential testing, while many modern hypotheticalists are *pure* consequentialists in holding that *only* testable consequences (and not generative information *per se*) can provide epistemic support.[2] For pure consequentialists, *all empirical support = empirical evidence (data) = successful test results = successful predictions = true empirical consequences of the claim tested (plus auxiliary assumptions)*.[3]

Popperians are pure consequentialists, but they go further still, to assert that only *novel* predictions are capable of providing support, in an appropriate sense of 'novel'.[1] We may call this view pure, *novel* consequentialism. The equation of epistemic support with successful test results of novel predictions we may term the 'central dogma' of Popper's epistemology. The Lakatosians accept this central dogma.

---

[1] In a curious sense, therefore, Popper and Lakatos *do* assign special probative weight to heuristically-employed information. Context of discovery is epistemically relevant in the negative sense that any information employed there cannot be used to support the resulting theory. See my [1985], where I reject all forms of 'the *per se* thesis'—that discovery moves *per se* have special justificatory weight.

[2] Methodologically, an infallibilist or strong epistemic optimist must be a generativist. Because of the fallacy of affirming the consequent, consequentialism cannot logically support strong optimism. But *pace* Popper and Lakatos, the converse is not true. Generativism does *not* entail infallibilism or strong optimism.

[3] Actually, a further distinction is necessary, between those consequentialists (and generativists) who restrict supporting information to empirical data and those who allow that satisfaction of theoretical constraints can furnish epistemic support. I hold the latter position while Popperians and Lakatosians seem to hold the former. See n. 2, p. 17.

Pure, novel consequentialism is a variety of *partial,* as distinct from *full* consequentialism. Partial consequentialists hold that only a proper subset of the empirical consequences of a theory are candidates for support. For example, Lakatosians follow Popper in excluding those consequences already known and used to construct the theory: 'of the empirically accepted logical consequences of a theory those . . . used in the construction of the theory fail to count in its support' (Worrall [1978a], p. 48). Clearly, Popper, the Lakatosians, and all pure, novel consequentialists are partial consequentialists. By contrast, John Stuart Mill [1843] and J. M. Keynes [1921] were *full* consequentialists (or virtually so). They regarded all empirical consequences as potentially supporting data, whether or not those consequences were known in advance and whether or not they were used to construct the theory. Notice, however, that a full consequentialist need not assign all consequences *equal* weight. (Mill's famous remark that in some cases a single instance provides a 'complete induction' while in others a myriad of instances fails to, suggests such a view.[2] One could even be a full consequentialist and also a weaker sort of novel consequentialist in the sense of giving novel consequences extra weight. (Some of Whewell's remarks are close to this.) More importantly for our purposes, *full consequentialism must not be confused with generativism.* Generativism is a constructive view of support. Consequentialism is critical-eliminative.

The epistemic idea behind generativism is simple: to derive a theory from what scientists already 'know'—from previous results which are solid and uncontroversial—is to provide support for the theory, support which in some cases (theoretical proofs) is the strongest form of theory justification we have. In the ideal case, a theory will be determined completely and uniquely by what we already know (but not so as simply to reformulate what we already know). To the extent that this ideal state of affairs is realised, the theory clearly has epistemic warrant, which means that the data and theoretical constraints employed in its derivation provide empirical support for it *prior to any empirical testing and regardless of whether the theory yields any new predictions*. To the extent that the ideal is attained, predictive testing becomes less important to scientific progress.

The best known historical example of a generativist methodology is Newton's method of 'deduction from the phenomena'. In its Newtonian form, generativism is near-infallibilistic and is therefore admittedly justificationist. But milder forms of generative, epistemic warrant are possible, just as milder forms of consequential warrant are available. Incidentally, the idea that heuristic derivability can make predictive testing less urgent is not merely a relic of the ancient past. Despite Lakatos's dislike of Newtonian methodology (see, *e.g.*, his [1978a]), he himself held that the stronger the heuristic of a research programme, the less important predictive testing and refutation become. Some of Lakatos's successors retain this view. Yet, unlike the older generative tradition, they provide no epistemic rationale for this anti-Popperian move.

---

[1] Several kinds of novelty have been distinguished in the recent literature. See, *e.g.*, Campbell and Vinci [1983].

[2] Mill [1843], Book III, Chap. 3, iii. See Grünbaum [1978] and Watkins's reply [1978] for entry into the wider issues that arise here.

Now most of the early investigators, including Newton, realised that the ideal of deduction from certifiably true data premises and guaranteed theoretical premises never can be achieved. Future exceptions to what we now consider knowledge might always turn up. For that reason, consequential testing retained an important place within classical generative methodology. However, these scientists and methodologists would have considered it strange that an element of fallibility should reduce the generative support to *zero*, so that empirical support now derives exclusively from empirical testing. It is true that a healthy dose of fallibilism makes one sceptical that sound discovery arguments can be found and places a premium on testing our ideas for the 'mistakes' presumably present, but fallibilism dulls the blade of consequential testing also. For consequentialists, too, must employ complex arguments; and these arguments typically contain many theoretical auxiliary assumptions of two kinds—those necessary to derive the predictions in the first place and those used to process and interpret the raw data. It is not only generativists who must rely on a body of previously 'established' results, commonly termed 'background knowledge'.

Deep fallibilism also casts suspicion on the Popperian–Lakatosian emphasis on boldness and novelty. Since truth often is very difficult to ascertain in science, is it not self-defeating to demand new content at every turn? To be sure, science aims at interesting truth, not just any old truth. But there are other forms of epistemic progress than making claims with 'excess content' over previous claims which themselves are rank conjectures.

Generativists marvel at how modern-day, pure consequentialists claim to have achieved a nearly complete reversal in the logic and methodology of science without recognising the boomerang of their anti-generativist arguments. The very information which formerly was capable of providing substantial support for theories is now said to provide zero support, always and everywhere; whereas the predictive tests that formerly supplemented generative justification are now considered the sole source of empirical support (*cf.* Laudan [1980]). I call the problems of historically explaining and logically justifying this reversal the mystery of the Great Logical Inversion (Nickles [1984] and [1987b]). The failure of Newtonian methodology does not automatically justify the Popperian central dogma.

5 THE 'PARAMETER FIXING' ARGUMENT

The difference in intuitions about empirical support can be brought out quite sharply by considering the use of data to fix a free parameter of an equation or theory. This is the Lakatosians' strongest supporting argument for the view that the same information cannot be used twice, upon pain of falling into *ad hoc* accommodation of theory to data. For example, Worrall ([1978a], p. 48) and Zahar ([1983], p. 245) agree that it is intuitively obvious that the data provide no support at all for the parameter-fixed equation or theory, *i.e.*, the equation with the parameter now replaced by its numerical value. Yet by seventeenth-century lights, it is intuitively obvious that there *is* (or may be) support; for the empirical determination of a free parameter amounts to a *partial deduction from the phenomena.* The derivation is partial because we took the initial equation or theory to be given and not itself heuristically derived.

Consider Zahar's recent discussion of the 'adaptation of theories to facts'. Zahar says some interesting things about heuristics, but one key argument fails.

> [S]uppose we are given . . . *n* points which are empirically determined . . . and . . . the assumption that the points are connected by a polynomial function of degree $(n-1)$. . . . It is obvious that a polynomial having the required properties can always be computed; it is also intuitively clear that the law thus constructed receives no empirical support from the *n* given points. Facts can therefore play a crucial positive role in the evaluation of certain parameters . . . [but] cannot be regarded as corroborating the theory precisely because they are implicated in its construction ([1983], p. 245).
>
> . . . Popper is perfectly right in distinguishing between the context of discovery and that of justification. In the methodological appraisal of theories it is not the heuristic devices as such, but certain logical considerations which play a significant role. In the considered examples it can be *mathematically proved* that the presence of sufficiently many parameters guarantees that a finite number of facts will necessarily be accommodated; for there are at least as many unknowns as there are equations. Since a *mathematical theorem* tells us nothing about *physical reality*, it follows that the facts which help us towards constructing a theory cannot simultaneously support it (p. 246; Zahar's emphasis).

This is not at all clear to me, nor would it have been to practitioners of the classical heuristic tradition. Everyone will agree that the *general assumption* of the first quoted passage—that a polynomial of degree $(n-1)$ can be constructed—is not supported by the *n* data points, for the reason Zahar gives in the second passage. But Zahar is not entitled to conclude that the *specific*, parameter-fixed law or theory that results is unsupported by these data. For the specific law, unlike the general assumption, is an empirical claim. Moreover, it seems to me that Zahar's argument is irrelevant to the issue of whether all empirical support must result from consequential testing rather than from mode of construction, for a result guaranteed by mathematics alone is no more empirically testable than it is generatively supported by the *n* data points. Therefore, Zahar's example provides no support for his general conclusions. Indeed, we may turn the tables on Zahar and the Lakatosians: far from being a paradigm case of zero support, as the novel consequentialists claim, empirical parameter fixing is an exemplary case of what I am calling generative support.

Suppose that scientists *A*, *B*, and *C*, starting from the same form of equation, set out to determine the value of a parameter of the equation. *A* uses available data to determine the value; *B* simply guesses the same value that *A* determined; and *C* also guesses, but incorrectly as it turns out. Now the Lakatosians must say that, prior to empirical test, all three procedures provide equal support, namely zero. After empirical test they must say that the first and third cases provide equal support, again zero (*cf.* Koertge [1978], p. 268). Yet *post hoc* 'testing' of the guess at the parameter value will, in some cases, consist of essentially the same procedure of parameter determination that *A* used in the first place, whence it is consequential testing (as opposed to generative determination) only in a trivial sense. In order to refute or corroborate the guess, the correct value must be determined from the data.

To be sure, consequential testing can test the form of the equation as well as the parameter value, whether or not that form was simply guessed. (Testing cannot do this in Zahar's polynomial case, however, since the mathematics alone guarantees the result.) But surely the data which *A* uses to determine the parameter value, provide some support for the parameter-fixed law candidate. For the empirical determination procedure selects this law claim from the infinity of other candidates which do not fit this data. Surely Zahar does not wish to say that this selection is empirically unfounded. Of course, if the higher constraints are nonexistent, or so weak that anything at all is compatible with the data, then there is no support because the data have not really *determined* anything.[1] I return to the idea of discovery as selection in Section **6**.

Full rational motivation for introducing equations with additional parameters consists of deducing the form of the equation from results previously established to be reliable. When this ideal is unattainable, a less conclusive sort of derivation (based, *e.g.*, on plausible theoretical hypotheses) is preferable to arbitrary guesswork. But even when an equation is simply guessed and the parameters determined empirically, it is still correct to say that the determining data support that equation in so far as those parameter values are concerned, *i.e.*, *relative* to the assumptions as to form. Again, generative support can be specific in this way, rather than total and unqualified.

The idea of *partial generative support* therefore enables us to answer the charge against Newtonian programmes that since a deduction of lawful theoretical claims from purely phenomenal premises is clearly impossible, such generative methodologies are completely untenable and totally useless. For the idea of partial generative support shows this 'all or nothing' objection to rest on a false dilemma. 'But then giving partial generative support such an important role just trivializes the generative programme,' the objector will counter. My reply is that the consequentialist faces much the same objection. Duhem showed, in effect, that one can achieve only an 'indirect' or 'partial' falsification—*relative* to the auxiliary assumptions which we suppose true (and similarly for failure to falsify). Now no one takes Duhem to have shown that consequential support (or disconformation) is trivial or useless. Why not take partial generative support equally seriously?

---

[1] Howson [1984] provides examples (from statistical estimation theory) more precise than my own in an attack on what I term 'pure consequentialism'. I thank a referee for calling Howson's paper to my attention. I shall discuss these matters at greater length elsewhere—and also the applicability of a Bayesian perspective.

To sum up: (1) Generative support (consequential support, too, for that matter) can be partial or specific, determining (or testing) part of a claim rather than the entire claim.[1] *Relative* to the assumptions made or conclusions drawn about the form of the equation, the data often determine the values completely. This is as far from zero support as you can get! How could the support for those parameter values be any stronger, *modulo* the fallibility of our starting assumptions and data-gathering? (2) What supports part of a claim provides some (lesser) degree of support for the claim itself. Thus the parameter-fixed equation is supported to some extent, although far from totally, by the data which fix the parameter(s). Consequential testing primarily tests those other assumptions or conclusions, not the parameter values. However, such testing may also call the data into question and hence the parameter values. The fact that theory tests can be used to 'correct' or 'clean up' the data does not undermine my point. Generativists are as entitled to be fallibilists with respect to the data as consequentialists are.

The usual view is that some data are used to set parameter values, then other data are used to test the result. It is not my purpose to deny testing an important role in science. Indeed, I reserve for testing a more important role than some Lakatosians do. A fuller discussion would treat a variety of more realistic examples than the one case I have discussed.

6 THE INCONGRUITY OF THE LAKATOSIAN *ad hoc*-NESS RULES

The obvious objection to my position is that if I count parameter-fixing data as genuinely supportive, then I must also permit blatantly *ad hoc* tinkering with an empirically defective theory so as to accommodate it to the data. For such tinkering amounts to localised theory-construction.

My reply is that such tinkering is sometimes acceptable, sometimes not. There are indeed situations in which I reject tinkering as bad science

---

[1] Glymour [1980] makes a similar point about confirmation. Yet on p. 114 he appears to accept the conclusion of the parameter fixing argument.

(whence it becomes 'tampering'). But what makes the work bad is *not* 'us[ing] the same fact[s] twice: once in the construction of a theory and then again in its support' (Worrall [1978a], p. 48). What is methodologically unacceptable is the arbitrary introduction of new parameters and other unmotivated changes in the equations or assumed relationships of the variables. (However, the data can sometimes furnish reasons for introducing new parameters also.) Popperians and Lakatosians have *mislocated* the source of the difficulty and, accordingly, have mistargeted their objections. Worrall's 'no double use' rule is unfounded. It is not the accommodation of theory to data which is *ad hoc* but the arbitrary introduction of the new parameters, or the new equations, themselves. For such moves are objectionable ('*ad hoc*') *whether or not the tampering yields claims with novel content*. What we need in such cases is *more* heuristics, not less—or rather, more empirico-theoretical determination of the overall form of the functions in question. As long as the theory changes are heuristically motivated by strong constraints, I have no objection.[1] Guessing is an often necessary complement to warranted constraints, not a replacement for them.

This is a point which Zahar can well appreciate, since it is one on which he himself insists, but this point supports rather than undermines the double use of data. Conversely, if heuristic derivability furnishes no epistemic support (even in the weak, Popperian sense) and therefore has no epistemic rationale, I am puzzled why Zahar attaches so much importance to heuristic motivation of theories and theory-changes. An economic rationale for Lakatosian heuristics is not enough, as I shall now argue.

First, I shall point out an unresolved tension among the various Lakatosian *ad hoc*-ness rules, specifically, the tension between Popper's conception of *ad hoc*-ness, which Lakatosians retain, and the new conception of *ad hoc*-ness which derives from Lakatosian heuristics. Second, I shall argue that the third *ad hoc*-ness rule, which is most Lakatosian and least Popperian in motivation, requires an epistemic rationale.

Zahar explicitly articulates and further develops Lakatos's ideas on *ad hoc*-ness by formulating three distinct criteria of *ad hoc*-ness:

> A theory is said to be *ad* hoc_1 if it has no novel consequences as compared with its predecessor. It is *ad* hoc_2 if none of its novel predictions have been actually 'verified'; . . . [A theory is] *ad* hoc_3 if it is obtained from its predecessor through a modification of its auxiliary-hypotheses which does not accord with the spirit of the heuristic of the programme.[2]

My first claim is that these three criteria of *ad hoc*-ness are mutually incongruous in the sense that their underlying intellectual motivations stand in conflict. The first two, 'Popperian' criteria (or *ad hoc*-ness prohibitions) have epistemic rationales. The first demands highly testable theories with new content and the second demands corroboration of some of this excess content.

---

[1] Provided that these are epistemically relevant heuristics as opposed to those introduced purely for reasons of economy of research. Here we need to distinguish epistemic *ad hoc*-ness from economical *ad hoc*-ness. See *below* for my move from heuristics to discoverability.

[2] See Zahar [1973], p. 101. Lakatos distinguished three kinds of *ad hoc*-ness in his [1970], p. 175, nn. 2 and 3. Zahar alters Lakatos's second criterion.

Now Popperians think that fallibilism rules out any direct, constructive route to the truth and thus virtually reduces the quest for truth to the detection and elimination of error. Accordingly, they consider 'new' results which are *derived* from available data and theoretical constraints to be *ad hoc* in the sense that the logical content of the new results does not go beyond the content of the premises used to derive them.[1] *On this view, employment of constructive algorithms, strong heuristics, and 'deductions from the phenomena' are* prima facie ad hoc *procedures.* What saves scientific inquiry from *ad hoc*-ery is radical departures and bold guessing, a non-rational element of human creativity. Urbach suggests such a view (in [1978], pp. 108 ff.), without fully accepting it. This opposition between the generative and the pure, novel consequentialist conceptions of empirical support is surely the '*prima facie* inconsistency' between justificationist methodologies and *ad hoc*-ness methodologies to which Lakatos referred in the quotation in Section 2 *above.* But, curiously, Lakatos and Zahar immediately proceed to introduce a third *ad hoc*-ness prohibition which fits the generative tradition and is *prima facie* inconsistent with the Popperian viewpoint.[2]

At first glance, the third *ad hoc*-ness rule may appear to be unproblematically Popperian: it forbids the deliberate adjustment of theory to data. But a closer reading reveals that the rule prohibits *any* heuristically unmotivated theory change. Hence, blind guesses and Popperian conjectures turn out to be *ad* hoc_3, regardless of their novel predictive power! By contrast, theories or theory-changes which are determined by approved algorithms or strong heuristics are not at all *ad* hoc_3.

The underlying intuitions of the 'Popperian' and the 'Lakatosian' conceptions of *ad hoc*-ness run counter to one another.[3] Although my own methodological sympathies lie with Lakatos, it will be instructive to question the point of Lakatosian heuristics from a Popperian standpoint. Must not the third rule have a clear epistemic, and not merely an economic, rationale? I present three arguments for an affirmative answer to this question.

First, error-detecting methods are a crucial part of nonconstructive, epistemic methodologies (just as generative methods are basic to constructive, epistemic methodologies). Clearly, Popper's *ad hoc*-ness rules are intended to increase the detectability of error. But the constructive, heuristic strand of Lakatosian methodology actually *de-emphasises* the importance of empirical testing. Hence, this heuristic strategy undercuts Popper's epistemic strategy and is quite risky or worse—*unless it is the realisation of an alternative epistemic strategy.*

---

[1] When the derivation is deductively 'right'; otherwise, it is inductive, which for Popperians is equally bad.

[2] Popper himself sketched a third notion of *ad hoc*-ness in [1963a], p. 241, but one quite different from the third, Lakatosian criterion—although inspiring it perhaps. 'The new theory should proceed from some *simple, new, and powerful, unifying idea* about some connection or relation. . . .' A Lakatosian research programme, like a Kuhnian paradigm, discourages essential novelty, any radical departure from its core.

[3] Popperians could say that T_2 is not *ad hoc*, although deducible from premises $P_1, P_2, \ldots, P_m$, as long as T_2 yields novel predictions relative to its predecessor T_1 ($\neq P_i$, for any i).

To elaborate, Popper emphasises severe testing and falsification, and he encourages bold departures from any current view, however well established. For all its faults, Popper's strategy is not ruined by a bad choice of starting point. Lakatosians tell us that the ideal research programme is one which possesses a heuristic powerful enough to determine completely the logical and temporal sequence of theories, *with no external, empirical input whatever, and certainly with no attention to falsification or modification of the hard core.* Where Popper's methodology is insufficiently constructive and therefore too quickly responsive to new information, Lakatosian methodology has the opposite problem. It is too heuristic-driven, especially since the programme core is not epistemically warranted. The more powerful the heuristic of a programme, say the Lakatosians, the more entitled we are to hold the hard core immune to criticism. Hence, the power of the heuristic of a research programme in effect measures how Lakatosian *versus* how Popperian the methods of that programme are. Thus we seem to have all the disadvantages of a classical, generative methodology without the advantages. To deny that the negative and positive heuristics of a research programme are epistemically warranted and that they provide any support to the products of that programme leaves the entire MSRP without an adequate, epistemic rationale. Lakatosian *ad hoc*-ness$_3$ is itself *ad hoc*, at the metalevel!

Second, if the Lakatosians insist on following Popper on boldness, novelty, nonrational discovery, and guessing, how can they at the same time adhere to the *ad hoc*$_3$ rule? Why must any theory change be in accordance with, and guided by, the heuristic of the relevant research programme? For this requirement would exclude a bold new theory with all manner of novel consequences, unless it was generated by an extant research programme. Since we may suppose that such theories are actually proposed from time to time, the *ad hoc*-ness$_3$ prohibition is not only epistemically unjustified but also uneconomical. Potentially valuable ideas are thrown out because they do not meet the Lakatosian admissibility standard, which amounts to a kind of constructivist demarcation criterion not unlike classical inductivist criteria (*cf.* Lakatos [1970], p. 175). Notice in particular that any major change of research programme (including branching off into alternative research programmes) is *ad hoc*$_3$, yet Lakatos himself believed in a certain amount of proliferation.

Lakatosians will reply that *ad hoc*-ness$_3$ applies only to the auxiliary assumptions and not to the 'hard core' or the 'positive heuristic' of a research programme. 'But why not?', a Popperian can well retort. 'If the thrust of your methodology is to permit only rationally motivated ideas and changes, why uncritically protect the very principles which will govern your entire research effort?'

As I would put the question: Why should confining ourselves within the bounds of a single research programme (or small set of them) increase our chances for finding the truth—***unless the heuristic together with its theoretical and data-claim inputs provides epistemic support prior to testing?***

The Lakatosians may further reply that a good idea can always be made the basis of a new research programme. Hence, the research programme idea does not really cramp their Popperian style; it only recognises that scientific ideas are developed gradually, stage by stage, not instantly formulated and tested. The answer to this reply is that, while the reply removes one disadvantage of MSRP to Popperian methodology, it does so only by making clear how globally Popperian the MSRP methodological strategy really is.

My third argument develops a more general response to the Lakatosian reply. I shall argue on *economic* grounds that the Popperian (and hence the MSRP) methodological strategy is *epistemologically* bankrupt. I term this methodological predicament 'the economic catastrophe'.

1. Economy of research requires that discovery (original generation of theories) be coupled to justification (epistemic criteria of appraisal) in many or most, but not all, instances of generation.
2. Hence, on the whole (but not in every single case), generation must be constrained by justificatory considerations.
3. But epistemically constrained generation amounts to partial deduction from what we already 'know'.
4. Hence, epistemically constrained generation provides partial generative support for the generated products.
5. These epistemic constraints cannot all be employed, instead, as equally powerful *post hoc* tests, upon pain of logical regress.
6. Hence, an economically viable methodology must include an epistemically relevant constructive component, that is, a component which furnishes generative support.
7. (Corollary) In denying that the constructive component of MSRP is epistemic, Lakatosians are dooming MSRP to hopeless inefficiency.

I cannot defend this argument in detail here (see my [1985]). I have space only for a few motivating remarks, which I hope will deflect the charge that I am simply begging the question by invoking the concept of generative support.

In my judgement, neither Popper nor Lakatos squarely faces what is surely the central problem of methodology: to show that the methods advocated have a reasonable chance of achieving the stated goals of the enterprise—in this case, finding true, interesting theories about the universe. A fallibilistic theory of inquiry rejects the justificationist stipulation that 'knowing' requires 'knowing that we know'. However, fallibilism does not thereby remove the responsibility of showing that our methodologies are reasonable means to our stated ends, that they are superior to sheer, blind luck—the complete absence of method.

Of the infinite number of candidate theories possible for any domain, scientists will, over time, actually produce and consider only a finite set of candidates. In effect, the discovery process, whatever its character, filters out these few from the limitless set of potential theories. Given this strong selection, there surely must be some coupling between modes of generation and criteria of epistemic appraisal.

Charles Peirce attempted to solve the problem (in part) by supposing that the human mind was atuned to nature—a solution few people take seriously. Popper's official position is that there is no such epistemic coupling. It is accidental which theories are generated, and scientific progress is an accident for which there is no rational explanation.[1] Lakatosians have attempted to explain scientific progress by adding their constructive component to Popper's method. By itself, however, this constructive component does not solve the epistemic problem of methodology. Nor do Zahar's efforts to establish the rationality of Lakatosian heuristics extricate us from the methodological predicament (economic catastrophe). Zahar does see Lakatos's (praiseworthy) move from general to specific ('knowledge-based'?!) heuristics as a fortunate narrowing of our freedom to choose among possible hypotheses. But surely this narrowing is fortunate only as long as it is epistemically justified. The generative component must be (in the main) 'epistemically relevant'.

Restated in terms of research programmes, the methodological predicament is this. Why think that any theory generated by a particular, historically available research programme has a decent chance of being true? Indeed, why think that any theory generated by *any* of the research programmes we shall eventually produce has a decent chance of being true?

It is not necessary to delve into Lakatos's positions on falsification, appraisal, and advice to see that the methodological pluralism of MSRP, in which competing programmes are comparatively evaluated by *post hoc* testing, does not solve the original problem, any more than Popper's monotheoretic methodology solves it. Short of embracing a comprehensive, eliminative induction strategy, MSRP is basically a Popperian, conjectures-and-refutations methodology, only the conjectured units are research programmes rather than individual theories.

---

[1] Popper [1965]. See Rescher [1978], Chap 3, for an excellent discussion of the economic catastrophe and of Peirce's and Popper's responses to it. Unofficially, Popper's 'methodological correspondence principle' ('Any adequate theory must yield its predecessor, to a first approximation') suggests a heuristic route out of this pickle. My argument will be that heuristics is not enough. One may ask why Popperians and Lakatosians adopt the correspondence requirement, given that it carries no epistemic weight and that, despite its heuristic virtues, it is sometimes *un*economical, by ruling out interesting theories already 'on the table'.

7 IS LAKATOSIAN METHODOLOGY INCOHERENT?

In the last section I argued that Lakatosian heuristics *must* be epistemic if the Lakatosians are to address the central problem of methodology. Otherwise, MSRP is radically incomplete. In this section I provide some additional arguments for generativism by pointing out how the Lakatosians' own methodological principles, as I understand them, require heuristics to be epistemic. Since the Lakatosians deny heuristics any epistemic role, MSRP is incoherent.

1. As the Lakatos quotation in 2 indicates, MSRP provides 'nonjustificationist criteria for appraising scientific theories based on anti-*adhoc*-ness'. Yet *ad hoc*-ness itself is an epistemic notion. As Worrall ([1978b], p. 322) observes, every *ad hoc*-ness prohibition corresponds to a concept of empirical support. How can this be so if heuristics provides no epistemic support? Assuming that the Lakatosians do not disagree among themselves on this point, it follows that either their position is inconsistent or else they must admit that an epistemic but nonjustificationist heuristics (a generative methodology) is possible.

2. By comparison with H–D methodology, MSRP attaches more importance to strong heuristics and less to consequential testing in virtue of its greater heuristic power. But that can be so only if the heuristic component of MSRP has epistemic significance, else the Lakatosians gratuitously reduce the overall support available to theories. For if failed tests carry less weight than for Popper, then so must successful tests carry correspondingly less weight (see point 4 below). Also, one of the strongest forms of empirical support is to show that no competitor of the theory has survived falsification. But since Lakatosians do not take falsification very seriously, this kind of support is less important to them. If we assume that consequential support is often genuine, why, we may ask, would anyone reject a methodology which furnishes *more* epistemic support for theoretical claims for a methodology which provides less? Fallibilism, it would seem, puts a premium on any epistemic support we can find.

3. In denying that any information used in context of discovery can count in support of the theory, the Lakatosians actually *do* give heuristics a kind of negative epistemic significance. Heuristics, however warranted and however employed, *decreases* overall empirical support rather than increasing it! There are logical objections to this (see 5 below), and again we have an apparently gratuitous rejection of an important source of epistemic support. How can adding a purely nonepistemic 'generator' to H–D methodology have such radical epistemic consequences? Because of the 'no double use' rule.

4. There is an objectionable, *epistemic asymmetry* in MSRP. On the one hand, failure to be heuristically derivable is an epistemic black mark against a theory—the theory is *ad hoc*$_3$. On the other hand, a tight, heuristic derivation does not count in its favour, epistemically speaking. Thus, we are

not allowed to say that Kepler's laws provided any support for Newton's theory of gravitation; only Newton's empirical departures from Kepler provided support (a view which appears to confuse support with selective support).[1] Popperians and Lakatosians often have stressed that you cannot have epistemic success without the possibility of failure; but now it seems that you can have epistemic failure without a corresponding possibility of success.[2]

5. Denying that heuristic arguments can, in principle, provide support, amounts, in many cases, to denying that an argument from better established premises ever provides support for its less-well-established conclusion. For example, Zahar [1983] sketches Planck's proof that Special Relativity requires a certain modification of Newton's Second Law. The premises of the argument are all solid, scientific results plus a 'correspondence principle' based on a warranted continuity assumption. Zahar himself uses the term 'proof'. Why then does this argument provide no support for the modified second law?

Zahar faces a dilemma here. Either the premises of the derivation are epistemically warranted or not. If they are, then, given that the conclusion is epistemically weaker than the premises, the argument furnishes epistemic support—and Zahar is wrong to deny it. If the premises are epistemically arbitrary, then the heuristic derivation is of doubtful importance; and Lakatosians are wrong to *require* that all admissible theories be heuristically derivable from an epistemically arbitrary core.

Lakatosians will reply that my cure for their problems is worse than the disease. The idea of generative support (they will say) runs into trouble right from the start, because it begins a logical regress which ends in metaphysical assumptions, which are, by definition, untestable. Zahar [1983] is at pains to stress the metaphysical character of Lakatos's 'hard core'.

A *tu quoque* reply is that Zahar himself faces a parallel, 'rationality' regress problem in trying to show that Lakatosian heuristics can be rational. As for the epistemic regress to metaphysics, I answer: (1) some elements of Lakatos's hard core are metaphysical not because they are intrinsically untestable but because they are deliberately protected from falsification. A generative methodology need not make its starting points immune to criticism. (2) There is no reason why all or any of the premises in a generative argument need be metaphysical. For example, Zahar himself acknowledges that the premises of Planck's proof (mentioned *above*) are all solid results, plus a warranted continuity assumption. One could cite many other cases, such as Ehrenfest's proof in [1911] that Planck's *empirical* law, together with some uncontroversial principles of statistical mechanics, logically entails a discrete weight function for energies (see Nickles [1976]). That is, it entails the existence of energy quanta, a new quantum statistics (as physicists later realised), and so on. Scientists consider a solid theoretical derivation of a claim to provide additional support regardless of whether the result in question yields new predictions. (3) While not just anything will count as a warranted premise in a derivation, neither should we set the standards impossibly high. In the *Opticks* Newton himself was content to assume as axioms or principles 'what hath been generally agreed on'.[1] (4) Even when metaphysical and other empirically unwarranted assumptions are employed, we may continue to speak of the generative support of *other* premises. Not every premise need function in the same way, although it remains true that one weak premise weakens the argument. But even the presence of a highly plausible and uncontroversial metaphysical principle need not spoil the epistemic warrant of the derivation as a whole, particularly when such principles have survived rational criticism. Generative support is often local and partial. The derivation of a single, new constraint which cuts down the known search space amounts to partial generative support, even if the remaining pool of possible problem solutions are evaluated by consequential testing.

---

[1] See Franklin and Howson [1985], which appeared after this paper was submitted.

[2] My asymmetry argument is one argument for giving theoretical constraints (and not just data) epistemic weight. Besides, now that everyone rejects the possibility of a general observational/theoretical distinction, where is one to draw the line? Lakatos himself frequently pointed out the importance of theory in determining and interpreting data claims. Actually, the Lakatosian view that no information used to construct a theory could possibly count in its support is a sophisticated form of the old, linear theory of justification. See Nickles [1987].

8 EMPIRICAL SUPPORT, GENERATIVE AND CONSEQUENTIAL

In denying that any information or methodological devices employed in theory construction provide empirical support to a theory, Lakatosians are rejecting two traditional positions which must be carefully distinguished. One is the Mill–Keynes position that all correct predictions provide empirical support, whether or not this data was known in advance and whether or not it was used to construct the theory. Popperians and Lakatosians oppose Mill and Keynes on this issue, holding that antecedent information (at least that used in theory construction) provides zero support and that novel predictions carry nonzero weight.[2] The other position rejected by the Lakatosians is the classical generative position.

Among the multitude of issues which arise here (*e.g.*, how we may compare the strength of generative and consequential data), the point I wish to make is that there *is* a sense in which we cannot use the same information twice. We cannot *count* the same information twice as providing *both* generative *and* consequential support, in those cases in which we can reason, say, from data to theory and from theory to some of the same data. This, I claim, is the correct point to be made about 'used' data.[1] Perhaps this point is what lies behind the Lakatosians' prohibition on using the same information twice. If so, they confuse *using* information twice with *counting* it doubly in justification, with the ironic result that they end up not counting it at all.

---

[1] Newton [1952], p. 20. For the important methodological point, see also Peirce [1877], end of Section IV.

[2] By 'nonzero weight' I do not mean that this weight confers a nonzero probability on the theory tested, only that successful and unsuccessful predictions count for something rather than nothing.

For example, Worrall ([1978a], p. 48) correctly states that because the Michelson–Morley interferometer result was used to construct the Lorentz–Fitzgerald contraction hypothesis, it does not *confirm* the latter, as a result of consequential 'testing'. But contrary to Worrall, this fact does not entail that the interferometer result provides no *support* for the contraction hypothesis. To say this simply assumes that all support is consequential and begs the question against generative support.

A *pure* consequentialist cannot consistently be a generativist of any sort. Nor can a *full* consequentialist be a generativist, for in some cases that would mean counting the same information twice—as affording both generative and consequential support. But some *partial* consequentialists *can* be generativists, namely those which are 'impure'. Lakatosians adopt a pure, partial, novel consequentialism and reject both generativism and full consequentialism.[2] My point has been that simple generativism is compatible with (some varieties of) partial consequentialism and that this two-pronged conception of justification is necessary to a viable, epistemic methodology. In defending generative support, therefore, I am *not* siding with Mill, Keynes, *et alia* on full consequentialism. I reject *both* sides of the Mill–Keynes *vs.* Whewell–Popper–Lakatos debate.

9 HEURISTICS AND GENERATIVE SUPPORT

Not all heuristic derivations endow their conclusions with epistemic support, of course. Some heuristics may be purely economical in function. Some heuristics themselves are completely untrustworthy, and some heuristic derivations employ unwarranted premises. When the premises are better warranted than the conclusion originally was, few people would hesitate to deem the argument supportive, but it may be supportive even when some premises are not as well supported as the conclusion. When canvassing all available arguments for a claim, a 'consilience' of arguments, including weak arguments, provides stronger support than a divergence of arguments—even when it is the weak arguments which diverge from the stronger.

---

[1] This matter deserves a closer look. Is it possible that the prediction test a different feature of the theory than the data was used to determine? If so, the information could be counted twice.

[2] But for indications that Lakatosians are receptive to a kind of prior support, see Lakatos [1968], p. 376, where he surprisingly speaks of 'prior acceptance', which is prior to testing but 'usually not prior to evidence'. See also Worrall's [1978b] reply to Koertge. I suggest that where information can be employed either as generative or as consequential support, it is usually better to use it generatively, since generative arguments usually are epistemically more powerful than consequential arguments. I shall not defend this claim here, for there are a number of complex cases to consider.

Is this generative support necessarily lacking if the hypothesis or problem solution in question was not originally thought up by heuristic means? Is generative support an historical rather than a logical conception of support in the sense of Musgrave [1974]?

No. Original 'discovery' arguments *per se* carry no *special* epistemic weight. Generative support is support provided by arguments. Hence, it matters not whether the arguments lay behind the original articulation of the claim or were only produced much later. Thus, Popper, Lakatos, and Zahar are right in thinking that there is a distinction to be made between the historical discovery process and the logic of justification. They are correct that the justification of any particular claim is (or could have been) *post hoc*, coming after the theory in question was formulated. But they are wrong in concluding that all justification is consequential. Not all *post hoc* justification is consequential justification. Generative justification may be *temporally post hoc* to the original discovery or formulation of the claim, just as consequential justification is. But generative arguments are not *logically post hoc*, in the sense of 'consequential'.

Since it is the warrant of the finished deduction and its premises which carries the weight, I prefer to speak of generative justificatory arguments as 'generat*ability*' or 'discover*ability*' arguments rather than simply as 'generation' or 'discovery' arguments. Generative justification amounts to *potential* discovery: however the claim in fact came to be conceived, it *could have been* discovered in this logical or warranted-heuristic manner, had all the relevant information and techniques been available at the time. To say this is not to deny that finding deductive arguments can require great creativity. We are here discussing justification, not creation.

In the mature, fertile, scientific disciplines, a great deal of rational reconstruction typically occurs between the original formulation or the tentative solution of a problem and the scientific community's eventual 'final' interpretation and acceptance of a solution. Consider the so-called 'Kepler problem' in mechanics, from Kepler himself to Newton to Euler, to today's textbook. Another striking example is Planck's problem *circa* 1900. Within six or seven years, Einstein and Ehrenfest had radically reconstructed Planck's problem and his solution, thereby practically inventing quantum theory in the process (Kuhn [1978]). By 1915 they, Debye, and others, had produced still another layer of reconstruction and had provided far more powerful techniques of solution. By 1924 the spelling out of the Bose–Einstein statistics provided yet another deep reinterpretation of 'Planck's' problem and its solution. These reconstructed products bore scant resemblance to Planck's own, original problem, the tools he employed to solve it, and the character of his solution.

On my account, a successful prediction furnishes consequential, epistemic support provided that the information was not employed in the *discoverability* argument for the claim in question (whence the information provides generative support). Hence, the question of what information was used to construct the claim is not an *historical* question at all but a *logical* question concerning the best discoverability arguments currently available. Thus we can avoid Zahar's serious problems in determining what an historical agent actually thought, intended, knew, or used at a given time. We need not get inside anyone's head. I suggest that as more and better discoverability arguments become available, they *should* affect our overall evaluation of empirical support, including predictive support. We juggle our conceptual systems around until we achieve the most coherent and best supported combination of claims.

Is there any reason for someone who accepts the logical construal of generative support to maintain an historical view of *consequential* support? I am not convinced that there is one. At any rate, the logical view avoids the temporal asymmetry problems which face historical conceptions of support.

Faced with a choice of using some information to derive a claim from the heuristic core of a programme or using that information to empirically test a theory, the poor Lakatosian is pulled in opposite directions. On the one hand, his methodology demands derivability from the core; on the other hand, such use of the information renders it epistemically useless. The burden of my paper has been to argue that an epistemic heuristics can avoid this problem—although I certainly do not pretend to have developed an adequate confirmation theory.

I conclude the paper with a necessarily brief response to three objections.

Objection 1 You argued in **6** that discovery must be coupled to justification in order to avoid economic disaster, hence that a viable methodology must have a generative (epistemic constructive) component. But now you seem to agree with Popper that discovery and justification can be decoupled.

Reply There is no contradiction. I said in **6** that there must be a weak epistemic coupling—many or most cases of original generation must be epistemically constrained. Coupling may be absent in any individual case but not in all cases. In any case a claim can be evaluated without knowledge of its mode of original generation. This is because, on my view, epistemic considerations furnish the same generative support, whether they had a role in the original generation or appear only as *post hoc*, reconstructed, derivability arguments. Discovery arguments *per se* carry no special weight—but they do carry nonzero weight if they are good arguments. Hence, context of *post hoc* justification can 'absorb' any justificatory consideration which were in fact employed in context of discovery—whether or not the *post hoc* justifiers are aware of that employment. While discoverability arguments usually are stronger than original discovery arguments, there normally will be some logical overlap and continuity between the two. For example, some of the original constraints will remain, perhaps in refined form.

Objection 2 First you criticise the Lakatosians for not giving their heuristics an epistemic role, but don't you yourself ultimately replace heuristics by discoverability considerations, which may not be very heuristic at all?

Reply The term 'heuristics' is ambiguous. If by 'heuristic motivation' is meant the derivability of a result from a special body of claims constituting 'positive science' or the core of a research programme, then discoverability and generative support are heuristic. However, if 'heuristic' pertains only to the original, 'creative' discovery argument or refers to a thought-guiding character special to some arguments but not others, then discoverability arguments need not be heuristic. The Lakatosians want their heuristics to be both things at once. On the one hand, it is clear that they value heuristics as a fertile source of creative ideas. On the other hand, much of their talk about heuristic derivation is really talk about heuristic deriv*ability* or discoverability. It is easy to imagine a case in which a theory is first rejected as *ad hoc*$_3$, yet is later found to be derivable from the negative and positive heuristics of the research programme—thus answering the *ad hoc*-ness charge. Derivability and not original derivation is what counts. The *ad hoc*-ness$_3$ prohibition is a logical not an historical rule.

Objection 3 But in emphasising epistemic derivability, you are replacing Lakatos's care-free, liberal, progressive view of heuristics by a dead, proof-ridden conception, and you are replacing the heuristically powerful MSRP by a conservative, epistemic methodology limited to deriving claims from what we already think we know. In recommending that the role of the hard core of MSRP be taken over by a body of positive science and that Lakatos's heuristic, nonepistemic methodology be transformed into a generative (epistemic) methodology, you are gutting Lakatos's MSRP of its heuristic content. This move is highly ironical for one who earlier criticised the Lakatosians for staking everything on the hard core of the research programme, and ironical for someone allegedly interested in scientific discovery.

Reply The charge that I am trying to resurrect classical justificationist ideas seriously overestimates my epistemic ambitions. And the charge that I stifle creativity undervalues the place I give discovery. Once we clearly distinguish discoverability from discovery, we can, to a degree, have it both ways. We can afford to be more radical at the discovery stage than at the discoverability stage, thus avoiding a complete conflation of heuristic power with epistemic warrant. To be sure, both discovery and discoverability must be (differently) coupled to epistemic criteria of appraisal, upon pain of economic catastrophe and failure to solve the central problem of methodology; however, this does not at all prevent aspiring discoverers from employing metaphysical ideas or whatever other source of inspiration they find. Let me spell this out.

First, I observe that astounding logical implications can and have followed from bodies of positive science, so, at its worst, the generative position is not so conservative as to make scientific revolutions impossible. (Recall the implications of Planck's empirical law.) Second, my own generativism is less conservative than this. I permit all the hypothesising and consequential testing that anyone could want; nor do I replace Lakatos's discovery heuristics by a set of positive scientific claims. I only insist that generative justification is an additional, important form of justification largely overlooked since hypotheticalist philosophers took over the field a century and a half ago. I have nothing like *ad hoc*-ness$_3$ to rule out as inadmissible interesting new ideas which are not discoverable from the current body of positive science. In many contexts, however, a claim is not well justified unless at least partial generative arguments eventually become available from the reconstructed field. Yet there is no justificationism here.

Third, I have argued throughout that, despite frequent disclaimers, the Lakatosian heuristics must have an epistemic function. I have wondered how this is possible, given the official Lakatosian position that the negative and positive heuristics of MSRP are epistemically arbitrary. The Lakatosians are deeply torn between giving heuristics a strong epistemic role and leaving it with a purely economic role of churning out new ideas at high speed. Their Lakatosian, constructivist principles require giving heuristics both roles; but doing so is inconsistent with their Popperian, anti-constructivist principles. Lakatos once wrote,

> *[T]he basic contradiction both in Whewell's and in Duhem's philosophy of science is their conflation of heuristic power and proven truth. Popper separated the two* ([1970], p. 123, n. 5, Lakatos's emphasis).

The Popperian tendency of Lakatosians is to avoid conflation by denying heuristic power any epistemic role. However, their constructive, Lakatosian tendency is to (tacitly) assign the heuristic component a significant epistemic role—whence, the Lakatosians themselves are guilty of conflation (albeit not as flagrantly as the justificationists were). Ironically, if we do interpret MSRP as a generative methodology, MSRP amounts to an attempt to unite logic of discovery with logic of justification in something like the seventeenth-century fashion. Such a view links original discovery and justification *too* tightly—whence the conflation charge.[1] Perhaps such a view is not implausible for low-level, data driven enterprises such as much seventeenth-century science and much twentieth-century social and behavioural science. But it *is* implausible for theory-driven methodologies such as MSRP. In the last analysis, proponents of MSRP only half appreciate that scientific ideas develop, and problems are solved over time, in stages.

---

[1] Actually, even Popper links discovery and justification too tightly by holding that what happens in context of discovery, as such, is relevant to justification—in a negative sort of way. See n. 1, p. 5.

In a mature, rapidly advancing, theoretical discipline such as many branches of physics over the past century or two, it simply will not do to make discovery and justification into a single process or at least into simultaneous processes. *Because* such disciplines are so splendidly creative, much stronger means of generative and consequential justification are likely to be available after the problem, the data and other constraints, and the original solution techniques have been the vehicle to their own rational reconstruction and/or replacement. You can't do everything at once. Learning to learn takes time and is itself a heuristic, bootstrap process.[1] Initial discovery and 'final' discoverability rarely coincide.

I have argued that by distinguishing original discovery from discoverability, that is, by distinguishing two functions for heuristics (or two senses of 'heuristics'), we can have both economical and epistemic adequacy. Since both are required of a viable methodology, any tenable methodology of science must be a generative and not merely a constructive methodology.

University of Nevada—Reno

REFERENCES

CAMPBELL, R. and VINCI, T. [1983]: 'Novel Confirmation', *British Journal for the Philosophy of Science*, **34**, pp. 315–41.

EHRENFEST, P. [1911]: 'Welche Züge der Lichtquantenhypothese spielen in der Theorie der Wärmestrahlung eine wesentliche Rolle?', *Annalen der Physik*, **36**, pp. 91–118. Reprinted in M. J. Klein (*ed.*), *Paul Ehrenfest: Collected Scientific Papers*, pp. 185–212. Amsterdam: North Holland, 1959.

Franklin, A. and Howson, C. [1985]: 'Newton and Kepler, A Bayesian Approach', *Studies in History and Philosophy of Science* **16**, pp. 379–85.

GLYMOUR, C. [1980]: *Theory and Evidence*. Princeton: Princeton University Press.

GRÜNBAUM, A. [1978]: 'Popper vs. Inductivism', in Radnitzky and Andersson [1978], pp. 117–42.

HOWSON, C. [1984]: 'Bayesianism and Support by Novel Facts', *British Journal for the Philosophy of Science*, **35**, pp. 245–51.

KEYNES, J. M. [1921]: *A Treatise on Probability*. London: Macmillan.

KOERTGE, N. [1978]: 'Towards a New Theory of Scientific Inquiry', in Radnitzky and Andersson [1978], pp. 253–78.

KUHN, T. S. [1978]: *Black-Body Theory and the Quantum Discontinuity, 1894–1912*. Oxford: Clarendon Press.

LAKATOS, I. [1968]: 'Changes in the Problem of Inductive Logic', in *The Problem of Inductive Logic*, pp. 315–417. Amsterdam: North-Holland.

LAKATOS, I. [1970]: 'Falsification and the Methodology of Scientific Research Programmes', in I. Lakatos and A. Musgrave (*eds.*), *Criticism and the Growth of Knowledge*, pp. 91–195. Cambridge: Cambridge University Press.

LAKATOS, I. [1971]: 'History of Science and Its Rational Reconstructions', in *PSA 1970, Boston Studies in the Philosophy of Science*, Vol. 8. R. C. Buck and R. S. Cohen (*eds.*), pp. 91–135. Dordrecht: Reidel.

LAKATOS, I. [1976]: *Proofs and Refutations*. Cambridge: Cambridge University Press. Origi-

---

[1] On learning to learn, see Shapere [1980].

nally published as a series of articles in *British Journal for the Philosophy of Science*, **14**, (1963–4).

LAKATOS, I. [1978a]: 'Newton's Effect on Scientific Standards', in J. Worrall and G. Curry (*eds.*), *The Methodology of Scientific Research Programmes: Philosophical Papers*, Vol. I, pp. 193–222. Cambridge: Cambridge University Press.

LAKATOS, I. [1978b]: 'The Method of Analysis-Synthesis', in J. Worrall and G. Currie (*eds.*), *Mathematics, Science and Epistemology: Philosophical Papers*, Vol. 2, pp. 70–103. Cambridge: Cambridge University Press.

LAUDAN, L. [1980]: 'Why Was the Logic of Discovery Abandoned?', in Nickles [1980], pp. 173–83. Reprinted with changes in Laudan, *Science and Hypothesis*, pp. 181–91. Dordrecht: Reidel, 1981.

MILL, J. S. [1843]: *A System of Logic*. London: Longmans, Green.

MUSGRAVE, A. [1974]: 'Logical versus Historical Theories of Confirmation', *British Journal for the Philosophy of Science*, **25**, pp. 1–23.

MUSGRAVE, A. [1978]: 'Evidential Support, Falsification, Heuristics, and Anarchism', in Radnitzky and Andersson [1978], pp. 181–201.

NEWTON, I. [1952]: *Optics*, based on the 4th edition, London, 1730. New York: Dover.

NICKLES, T. [1976]: 'Theory Generalization, Problem Reduction, and the Unity of Science', in A. Michalos and R. S. Cohen (*eds.*), *PSA 1974* (*Boston Studies in the Philosophy of Science*, Vol. 32), pp. 41–74. Dordrecht: Reidel.

NICKLES, T. (*ed.*) [1980]: *Scientific Discovery, Logic and Rationality*. Dordrecht: Reidel.

NICKLES, T. [1984]: 'Scoperta e Mutamento Scientifico' ('Discovery and Scientific Change'), *Materiali Filosofici 10*, 7–27. (English version available upon request.)

NICKLES, T. [1985]: 'Beyond Divorce: Current Status of the Discovery Debate', *Philosophy of Science*, **52**, pp. 177–206.

NICKLES, T. [1987a]: ''Twixt Method and Madness', in N. Nersessian (*ed.*), *The Process of Science*, pp. 41–67. Dordrecht: Reidel.

NICKLES, T. [1987b]: 'From Natural Philosophy to Metaphilosophy of Science', in P. Achinstein and R. Kargon (*eds.*), *Theoretical Physics in the 100 Years since Kelvin's 'Baltimore Lectures'*, pp. 507–41. Cambridge: MIT Press.

PEIRCE, C. S. [1877]: 'The Fixation of Belief', *Popular Scientific Monthly*, **12**, pp. 1–15. Widely reprinted.

POPPER, K. [1963a]: 'Truth, Rationality, and the Growth of Scientific Knowledge', in *Conjectures and Refutations*, pp. 215–50. New York: Basic Books.

POPPER, K. [1963b]: 'Science: Problems, Aims, Responsibilities', *Federation Proceedings, Federation of American Societies for Experimental Biology*, **22**, Part I, pp. 961–72.

POPPER, K. [1965]: 'Of Clouds and Clocks', as reprinted in *Objective Knowledge*, pp. 206–55. Oxford: Oxford University Press, 1972.

RADNIZTKY, G. and ANDERSSON, G. (*eds.*) [1978]: *Progress and Rationality in Science*. Dordrecht: Reidel.

RESCHER, N. [1978]: *Peirce's Philosophy of Science*. Notre Dame, Indiana: Notre Dame University Press.

SHAPERE, D. [1980]: 'The Character of Scientific Change', in Nickles [1980], pp. 61–116. Reprinted in Shapere's *Reason and the Search for Knowledge*, pp. 205–60. Dordrecht: Reidel, 1984.

URBACH, P. [1978]: 'The Objective Promise of a Research Programme', In Radnitzky and Andersson [1978], pp. 99–113.

WATKINS, J. W. N. [1978]: 'Corroboration and the Problem of Content-Comparison', in Radnitzky and Andersson [1978], pp. 339–78.

WORRALL, J. [1978a]: 'The Ways in Which the Methodology of Scientific Research Programmes Improves on Popper's Methodology', in Radnitzky and Andersson [1978], pp. 45–70.

WORRALL, J. [1978b]: 'Research Programmes, Empirical Support, and the Duhem Problem: Replies to Criticism', in Radnitzky and Andersson [1978], pp. 321–38.

ZAHAR, E. [1973]: 'Why Did Einstein's Programme Supersede Lorentz's?', *British Journal for the Philosophy of Science*, **24**, pp. 95–123, 223–62.

ZAHAR, E. [1983]: 'Logic of Discovery or Psychology of Invention?', *British Journal for the Philosophy of Science*, **34**, pp. 243–61.

Brit. J. Phil. Sci. 38 (1987), 207–224 *Printed in Great Britain*

Popper, Prior Probabilities, and Inductive Inference

by COLIN HOWSON

1 *Introduction*
2 *The Classical Theory and Large Universes*
3 *Jeffreys's Simplicity Postulate*
4 *The Status of Prior Probability Distributions*
5 *Subjective Probability Distributions*

1 INTRODUCTION

Popper is celebrated for his thesis that if $\boldsymbol{h}$ is any non-tautologous universal hypothesis interpreted over an infinite domain D, and $\boldsymbol{e}$ is any statement describing the properties of a finite set of individuals in D, then

$$P(\boldsymbol{h}/\boldsymbol{e}) = 0 \tag{1}$$

for any 'logical' probability P (what he means by 'logical' is that P be defined without invoking any assumption which could be regarded as entailing a material statement about the structure of D). The majority of Popper's arguments for (1) are based on his use either of so-called 'classical' measures (for a brief account of which see the Appendix at the end of this paper) on certain types of underlying possibility-space, or on close relatives of them, namely independence measures on product spaces. In this paper I shall argue that not only have such measures no privileged status in supplying the foundation for a theory, even a null theory like Popper's, of inductive probability, but also, for reasons which, as I shall point out Popper himself makes quite clear, no *a priori* measure has. Popper's insistence on a favoured prior distribution, even if it has anti-inductivist implications, is therefore at odds with his generally fallibilist philosophy. The remainder of this paper will demonstrate that this, I claim more characteristic, Popperian approach is perfectly compatible with a defence of the claim that the probability calculus can nevertheless be interpreted as a, or the, logic of inductive inference (this last claim is, of course, something that Popper has always contested). This claim will, of necessity, rest on a subjective interpretation of the probabilities. The last section of the paper will be devoted to ridding the current theory of subjective probability of an unnecessary if not harmful philosophical accretion (its identification of these probabilities with degrees of belief), and to putting the Dutch Book argument on a more satisfactory foundation.

Before proceeding to the body of the paper a brief remark on formal matters is in order. The probability calculus I shall take to be Kolmogorov's axioms (Kolmogorov [1933]), minus the axiom of countable additivity, and including the 'definition' of conditional probability. This last will be regarded not as a definition but as a further axiom relating a function $P(\boldsymbol{a}/\boldsymbol{b})$ of two variables to a function of one variable. Kolmogorov's axioms state that the domain of the probability function(s) is a field or algebra of subsets of a set (*i.e.*, a set of sets closed under the Boolean operations). However, every Boolean algebra is isomorphic to a field of sets and the sentences of a standard language are, identifying equivalent sentences, a Boolean algebra. It follows that a function satisfying Kolmogorov's axioms whose domain is a field of sets corresponding to a given algebra of sentences induces a function defined on the sentences themselves, and it is this function we shall call the probability measure rather than its associated set function. I do not, by the way, want to be taken as asserting that one should always regard a probability function as defined on sentences, or even open formulas; for the inductive interpretation of probability it is appropriate to do so, but not necessarily otherwise.

2 THE CLASSICAL THEORY AND LARGE UNIVERSES

An obvious difficulty involved in denying (1) is due to the fact that there is an infinity of possible structures which an infinite universe might possess (and depending on the conceptual resources at our disposal this infinity may be very large indeed). Let us consider a simple example. Suppose we consider only such universes in which the possession or not of a property A by its members is all that interests us. Such a universe might be that which is determined by the outcomes of $\boldsymbol{n}$ tosses of a coin, where A is the property of being a head; or in general it might be that universe of possibilities determined by any sequence of $\boldsymbol{n}$ experiments whose outcomes are simple alternatives. There are $2^{\boldsymbol{n}}$ possible distributions of the property A among $\boldsymbol{n}$ individuals, and as $\boldsymbol{n}$ tends to infinity the number of universes in which all the individuals have A—*i.e.*, one—as a proportion of the whole tends very rapidly to 0, obviously. Moreover, $P(\forall \boldsymbol{x} A(\boldsymbol{x})/A(\boldsymbol{a}_1) \wedge \ldots \wedge A(\boldsymbol{a}_m)) = 2^{\boldsymbol{m}-\boldsymbol{n}} \rightarrow 0$ as $\boldsymbol{n} \rightarrow \infty$ $(\boldsymbol{m} < \boldsymbol{n})$. Thus a straightforward application of the classical theory leads to (1) by a simple limiting argument.

In this application the space S (see Appendix) is a set of state descriptions, or in model-theoretic language the set of all distinguishable models of the corresponding monadic predicate language with $\boldsymbol{n}$ individual names, and whose logical axioms are supplemented by a sentence asserting that there are exactly $\boldsymbol{n}$ individuals, each named by a constant. It is easy to see that (1) is obtained by the same method where there are no individual constants and no additional predicates (one can always define constants by means of monadic predicates and equality). The space S now has $n+1$ instead of 2^n members (for now only the numbers of A-individuals distinguish the models), and the probability of $\forall x A(x)$ given that m individuals are A is therefore $(n-m+1)^{-1}$. Although we have used particularly simple languages, the result (1) is typical: any consistent nontautologous law-sentence formulated in any first order language has zero probability in the limit as the (finite) cardinality of its domain tends to infinity (this follows from Theorem 4 of Fagin [1976]).

Another application of the classical theory dispenses with the limiting argument: if we actually start out with a countable infinity of named individuals then the universe satisfying $\forall x A(x)$ is a point in a continuum which can be identified with the binary expansions of all the real numbers in the closed unit interval. This has (Lebesgue) measure 1 and each point, like $\forall x A(x)$, measure 0, while the condition $(A(a_1) \wedge \ldots \wedge A(a_m))$ corresponds to a nondegenerate interval and so has positive measure. Again, the ratio is zero.

The argument above, or at any rate that part of it which uses the limiting procedure applied to the state description case, is actually used by Popper (Popper [1959], p. 365). He does, however, qualify that argument slightly by remarking that it is 'not entirely adequate, although it is . . . essentially correct' (*ibid.*). What is essentially correct in it, according to Popper, is that it makes the sentences $A(a_i)$, $\neg A(a_i)$ in any state-description probabilistically independent with constant probability (in this case, equal to one half). In other words we are required by Popper to assume that the a_i correspond formally to the outcomes of a sequence of Bernoulli trials with parameter p where $0 < p < 1$. Clearly, (1) is still obtained. My fundamental objection to Popper's strategy does not, as a matter of fact, depend on his having adopted one prior probability distribution as opposed to any other, as the later discussion will I hope make clear. However, it is of some significance to his own programme that those distributions he does choose have a model in the random selection of black and white balls from an urn with some given proportion of black and white balls. I shall return to consider the appropriateness of such an assumption later.

3 JEFFREYS'S SIMPLICITY POSTULATE

Popper makes unqualified use of the classical theory when he comes to discuss Jeffreys's Bayesian theory of inductive probability. This theory has been largely neglected by recent authors, and those few who do discuss it are apt (with the notable exception of Shimony [1970]) to convey a rather misleading impression of what Jeffreys actually asserts. I shall, therefore, briefly recapitulate Jeffreys's theory. In his [1921], written jointly with Dorothy Wrinch, and again in his [1948], Jeffreys had proposed that the prior probabilities of hypotheses vary in the same direction as their degree of simplicity. In the [1921] article he and Wrinch actually define an integer-valued measure, which I shall denote by π, of simplicity for the class of hypotheses expressed by differential equations with rational coefficients (though as any real number is approximated arbitrarily closely by rationals this is not a very restrictive condition). He was able to demonstrate that positive probabilities can be distributed over this denumerably infinite set in such a way that $\pi(\boldsymbol{h}_1) < \pi(\boldsymbol{h}_2)$ iff $P(\boldsymbol{h}_1) < P(\boldsymbol{h}_2)$. $\pi(\boldsymbol{h})$ is defined to be equal to the sum of the order, degree and the absolute magnitudes of the coefficients of $\boldsymbol{h}$ (fractions having been cleared), which means that in the integrated form of $\boldsymbol{h}$ $\pi(\boldsymbol{h})$ will reflect among other things the number of undetermined parameters in that form.

By the time of his [1948] Jeffreys had come to regard the restriction to differential equations as unsatisfactory and proposed a measure of π' of simplicity applicable to any given hypothesis: its degree of simplicity, or rather complexity, is now simply the number of its undetermined, or adjustable, parameters. He had also abandoned, though this is more implicit than explicit, the idea of a globally defined *a priori* measure, *i.e.*, one defined on the class of all possible hypotheses (one reason possibly being that this class is no longer well defined). Indeed, we often find him assigning prior probabilities of one half each to two competing hypotheses that do not by themselves create a logical partition (*op cit.* p. 220, for example): thus the ordering is imposed only on competing explanations of the same class of phenomena. Furthermore, though probability varies still with simplicity, it now seems to be required only that $\pi(\boldsymbol{h}_1) < \pi'(\boldsymbol{h}_2) \Rightarrow P(\boldsymbol{h}_1) \leqslant P(\boldsymbol{h}_2)$, equality being allowed apparently where background information indicates that there is as strong a case for admitting a new parameter as not. Thus on pages 101, 220 *op. cit.* he states that if $\boldsymbol{h}$ with $\boldsymbol{m}$ free parameters is obtained from $\boldsymbol{h}'$ with $\boldsymbol{m}+1$ by setting one of them equal to a fixed number, say zero, then if $\boldsymbol{h}$ and $\boldsymbol{h}'$ are the only two seriously proposed alternatives then $P(\boldsymbol{h}) = P(\boldsymbol{h}') = 1/2$ (in sharing out all the probability only among the seriously proposed hypotheses Jeffreys is being what Shimony calls a 'tempered personalist' (*op. cit.* p. 101)).

The purpose of the simplicity postulate is to rule out, among other things, the possibility of those scientific theories which are simply modified *ad hoc* to agree with experimental data achieving any considerable confirmation thereby. This sort of *ad hoc*kery is usually exhibited in the case of mathematical hypotheses by the introduction of as many new parameters into the theory as are required to achieve consistency with the data. Consider, for example, two hypotheses $\boldsymbol{h}$ and $\boldsymbol{h}'$ where $\boldsymbol{h}$ predicts the data directly but $\boldsymbol{h}'$ only at the cost of admitting $\boldsymbol{m}$ new parameters. For convenience let $\boldsymbol{m} = 1$ and the parameter be $\boldsymbol{a}$. It is required that, other things being equal, $\boldsymbol{h}$ should obtain a vastly greater confirmation from the data than $\boldsymbol{h}'(\boldsymbol{a}')$, where $\boldsymbol{a}'$ is the value of $\boldsymbol{a}$ computed from the data, and that this should be the case independently of the actual value of $\boldsymbol{a}'$. This means, in Bayesian terms, that the prior probability of every instance of $\boldsymbol{h}'(\boldsymbol{a}')$ over all permitted data points $\boldsymbol{a}'$ should be negligible compared with that of $\boldsymbol{h}$

(assuming that the prior probability of $\boldsymbol{h}$ itself is not close to 1). Moreover, the sum or integral $\int P(\boldsymbol{h}'(\boldsymbol{a}'))\mathrm{d}\boldsymbol{a}'$ over the set of values $\boldsymbol{a}'$ should not exceed $P(\boldsymbol{h})$ either; were it to do so then that would amount to accepting that the data discriminate in favour of $\boldsymbol{h}$ for some values of $\boldsymbol{a}$, over $\boldsymbol{h}$. Thus we obtain $P(\boldsymbol{h}) > P(\exists \boldsymbol{a} \boldsymbol{h}'(\boldsymbol{a}))$ consistently with the simplicity postulate.

Now hypotheses with adjustable parameters are Σ-sentences, *i.e.*, they are sentences prefixed by a string of existential quantifiers. This being the case it is difficult to see why Popper should want to assimilate them to universal generalisations in satisfying (1). Nevertheless he does, and in so doing has to give reasons for finding the simplicity postulate illegitimate. His argument, which is contained in his [1959] Appendix* viii, is explicitly based on the classical measure. I shall present it with the aid of a simple example. Suppose that $\boldsymbol{h}$ has one adjustable real-valued parameter $\boldsymbol{a}$ and that $\boldsymbol{h}'$ has two, $\boldsymbol{a}$ and $\boldsymbol{b}$. Suppose also that $\boldsymbol{b}$ takes the value 0 on $\boldsymbol{h}$, so that $\boldsymbol{h}$ can be identified with the line $\boldsymbol{b} = 0$. Finally, for the sake of simplicity, suppose that the values permitted $\boldsymbol{a}$ and $\boldsymbol{b}$ by $\boldsymbol{h}'$ exhaust a bounded region R of the $(\boldsymbol{a}, \boldsymbol{b})$ plane with the exception of the points $\boldsymbol{a}$ on the line $\boldsymbol{b} = 0$ in R, which correspond to the values of $\boldsymbol{a}$ permitted by $\boldsymbol{h}$. It follows that if the probabilities of the hypotheses are set proportional to the Lebesgue measure of these two sets of points then $\boldsymbol{h}$ has probability zero. At any rate the probability of $\boldsymbol{h}$ does not exceed that of $\boldsymbol{h}'$. Both probabilities may of course be zero, depending on whether their associated parameter spaces are embedded in a higher dimensional set of points. Indeed, it is simple to construct another hypotheses $\boldsymbol{h}''$ with three free parameters $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{c}$ such that $\boldsymbol{h}'$ is obtained from $\boldsymbol{h}''$ by setting $\boldsymbol{c}$ equal to 0. Proceeding thus we can construct a denumerable sequence of mutually exclusive hypotheses of increasing complexity, and of non-decreasing probability. It follows that all those probabilities must be zero. Though the argument assumed that the possible values of $\boldsymbol{n}$ free parameters are confined within a bounded region R of $\boldsymbol{n}$ dimensional space the ratios of the probabilities of pairs of successive hypotheses obtained as above is always $0 : \boldsymbol{p}$ and is therefore independent of R.

It is remarkable that Popper should have advanced an argument of this type against Jeffreys, since the latter devotes much of his discussion to the defects, from his point of view, of prior assignments of the type Popper rests his objections on; and it is precisely because such assignments lead to (1) quite generally that Jeffreys rejects them. Nothing of this emerges in Popper's discussion, which he concludes with the claim that Jeffreys's theory contradicts the probability calculus. If it does then it is certainly not for the reasons Popper advances. Watkins in his [1985], pages 113–16 repeats the charge and shows that if the simplicity postulate were the thesis that hypotheses with more adjustable parameters are strictly more probable, and those with the same number are equally probable, without qualification, then the postulate is inconsistent; however, despite some obscurity in Jeffreys's exposition it is certain that he would disown this interpretation.

In the first place it is just very obviously inconsistent (there are, for instance, denumerably infinitely many hypotheses of the form $f(x) = kx$ with rational k, which cannot, obviously, have the same positive probability), and somebody of Jeffreys's logical and mathematical acumen would hardly fail to notice this. Moreover, such a principle is flatly contrary to Jeffreys's purpose of depreciating the confirmation of hypotheses constructed *ad hoc* from the data. In the example above, h and $h'(a')$ have no adjustable parameters and the latter is required to possess a negligible prior probability compared to the former. Jeffreys in fact assigns a continuous distribution of prior probability over the values of a where these possible values exhaust a nondegenerate interval of reals, which means that the prior probability of $h'(a')$ is actually zero (though its probability density is positive) whereas $P(h)$ is not only not zero but considerable (*cf.* Jeffreys *op. cit.* p. 101 *ff*).

Popper has other arguments, not relying on the classical theory of probability, which purport to exhibit inconsistencies in Jeffrey's theory. They are discussed in Howson [1973], and they seem to be no more successful than the above. Indeed, it is Popper's own position which is open to doubts as to its consistency. In the first place, no uniformly distributed measure over an infinite or semi-infinite interval of Euclidean n-dimensional space can be normalised, so there can be no uniform probability distribution over the unrestricted parameter space of a theory with n free parameters. Secondly, even if the parameter values are restricted to a bounded region, any nonlinear reparametrisation will destroy the uniformity of the measure, which is the main reason why historically uniform prior measures ceased to be regarded as an integral part of Bayesian methodology.

4 THE STATUS OF PRIOR PROBABILITY DISTRIBIUTIONS

However, these technical objections do not seem to me really to get to the heart of the matter, which is the question why prior probabilities should be assigned in the way Popper urges. There is certainly nothing in logic that tells us that this must be so. Now suppose that for any given method of assigning probabilities *a priori* we think of the sort of physical probability model or models which yield the same values, assuming that we are permitted to think of types of world as outcomes of some stochastic trial. We do not have to look very far to see what sort of model gives the values obtained *via* that 'classical' method which assigns equal probabilities to the 2^n $2-Q$-predicate state descriptions of length n. It is just that in which the elementary possibilities, state descriptions or points in a continuum, are completely randomly generated and in which therefore very strong conditions of independence hold (this is why Carnap rejected the measure $m\dagger$).

It is certainly legitimate to question why the probability function characterising such a random model, or more generally any model generating independence in a product space, should be thought the only one appropriate for assigning probabilities *a priori*. Why should we assume that the correct evaluation *a priori* of, say, the probability that the ($i+1$)st individual examined will be A conditional on the first i being A be that which characterises the picking of balls at random from a randomly selected urn? (For a detailed discussion of the urn models underlying various inductive-probability measures see Kuipers [1978]). Why indeed? For random models of possibility spaces in which n determinate individuals are distributed among k cells of a partition are maximally disorderly, and a prior assumption of randomness here is equivalent to an assumption that generalisations, *i.e.*, very low entropy states are extremely unlikely to hold in the limit as the universe becomes very large without bound. In other words to adopt the assumption of Popper's arguments is to commence with an extreme bias against the possibility that generalisations of any type will hold. If we split possibility space up into the two categories 'some generalisations hold' and 'no generalisation holds' then Popper's prior probability distribution assigns probability one to the latter in the limit (as also do, incidentally, all the regular c-functions in Carnap's λ-continuum of inductive methods, *cf.* Niiniluoto [1976], p. 436). This casts doubt, to say the least, upon the claim sometimes made (and made by me, I regret to say, in my [1984a]) that uniform prior distributions correspond to an attitude of epistemic neutrality between possibilities. *It all depends on how you characterise the possibilities.* The choice of ultimate partition is not a logical matter (*cf.* Miller [1974], p. 177); indeed it is determined entirely by what you think is an appropriate system of categories, which itself is something which will contain a considerable *a posteriori* element. To assume *a priori* that the possibilities you wish to be neutral between are those determined by the distribution of individuals among the cells of a partition, possibly a continuous one, is both arbitrary and, given the evidence which strongly suggests that the universe is a highly structured place, perverse.

It is hardly surprising that, having (completely *a priori*) adopted a position maximally inimical to the possibility of establishing general laws, Popper should be unable to offer any sort of positive solution to the problem of induction. For indeed he cannot, despite his claims to the contrary. This can be seen very easily in the context of a simple example. Let us suppose that $P(\boldsymbol{h}) = 0$ for all synthetic universal hypotheses $\boldsymbol{h}$. Suppose also that we have two competing explanatory hypotheses $\boldsymbol{h}_1$ and $\boldsymbol{h}_2$ which disagree about the outcome of an experiment whose actually observed outcome is $\boldsymbol{e}$, and that $\boldsymbol{h}_1$ implies $\boldsymbol{e}$ (modulo initial conditions *etc.*). Now it is easy to see that $P(\neg\boldsymbol{h}_1) = P(\neg\boldsymbol{h}_2) = 1 = P(\neg\boldsymbol{h}_1 \wedge \neg\boldsymbol{h}_2)$ given Popper's thesis about the probability of laws. In other words $\neg\boldsymbol{h}_1$ and $\neg\boldsymbol{h}_2$ are independent 'events', and $\neg\boldsymbol{h}_2$ has actually occurred ($\boldsymbol{h}_2$ has been falsified by the experiment). Since P is by assumption nonarbitrary we must therefore accept, on pain of committing the gambler's fallacy, that the falsification of $\boldsymbol{h}_2$ makes it no less likely that $\boldsymbol{h}_1$ is false now than before the experiment. Thus a subsequent appeal to $\boldsymbol{h}_1$'s surviving the test, *however severe that test*, in any context in which the truth values of its consequences are of either practical or theoretical significance, is quite illegitimate. Popper claims ([1972], ch. 1) that one may infer that $\boldsymbol{h}_1$ is nevertheless preferable to $\boldsymbol{h}_2$ simply on the ground that it has not yet been falsified. But even were one to grant him this—and I argue in my [1984a] that one should not—it is still no solution to anything like the classical problem of induction, because there is nothing forthcoming from the experiment about the truth or falsity of $\boldsymbol{h}_1$: we are no wiser than before about the identity of the true theory.

This result appears to put Popper at a disadvantage compared to the Bayesian, who explicitly disavows (1), at any rate as a general rule, and who is able to infer from a sufficiently severe but successful test to a considerable posterior probability of the surviving hypothesis (a severe test of $\boldsymbol{h}$ in Bayesian terms is one which has an outcome $\boldsymbol{e}$ predicted by or merely probable given $\boldsymbol{h}$ but very improbable relative to all the other plausible alternatives to $\boldsymbol{h}$; on such an outcome the term $P(\boldsymbol{e}/\neg\boldsymbol{h})$ in the denominator of Bayes's Theorem is small). Since, moreover, the consequences of $\boldsymbol{h}$ have at least as large a probability as $\boldsymbol{h}$ there is great practical significance in possessing a very well supported theory: its predictions can be relied on to at least the same extent as the theory itself.

But notoriously these benefits are bought at the price of having to justify a prior probability distribution which confers positive probabilities on a sufficiently extensive class of hypotheses. But a prior distribution not satisfying (1) is no less (but also no more) arbitrary than one which does, in so far as its intended function is to supply a standard for judging the correctness of inductive inferences—in so far as it is endowed with a logical function, in other words. There is no need to argue this further if the burden of the previous discussion of Popper's attempts to establish (1) is accepted. Ironically, Popper no less than the inductive logicians whom he takes to task for so doing, has overlooked his own injunction to express explicitly as a hypothesis any assumption which is not a truth of logic or a contradiction (see *e.g.*, Popper [1959], p. 368), and it is therefore strange that he, of all people, failed to recognise that *any* prior distribution over a class of hypotheses, which is endowed with objective significance ought itself to be regarded as a hypothesis—as a type of statistical hypothesis, in fact. To assimilate such hypotheses to the status in effect of logical axioms is to deny fairly substantial assumptions the opportunity of any sort of empirical control. Even were they classified explicitly as hypotheses most of them would still escape the possibility of empirical test, since they are purely metaphysical. When, for example, the prior probability of Special Relativistic mechanics is compared to that of Newtonian then the space of possibilities implicitly under discussion is a class of possible types of large-scale universe. Or it may be a probability distribution over a space of parameter values that is involved. One cannot sample such populations (indeed, the idea of so doing is practically meaningless) and so one cannot bring any data to bear on the question of whether these or any derivative hypotheses are true about them or not.

Now interestingly enough Popper's approved prior distribution over state descriptions with two 'observable' Q-predicates A, $\neg A$ (namely that determined by the conditions of independence with constant probability p, $0 < p < 1$, of A) is, if formulated as a statistical hypothesis, one of those which in fact are testable, and testable by the usual methods for testing any binomial hypothesis. Moreover, it will be rejected at practically any level of significance in a considerable number of cases. Those cases are, of course, just those in which there is a very highly confirmed hypothesis that in such experimental situations a particular effect is invariable forthcoming. Experimental science knows of many such experiments. Bizarrely, Popper even tells us how, in such cases, we may obtain overwhelming inductive-probabilistic support for the hypothesis that the parameter p lies in the interval $1-\delta$ for arbitrarily small $\delta > 0$. For in Appendix* ix [1959], p. 417, he informs us that we may take the prior probability for a binomial parameter to be uniformly distributed over the unit interval. It follows, as Laplace was the first to prove, that the posterior probability that the true value of the parameter lies in an interval including the observed frequency tends to 1 (in fact the posterior distribution for p is a beta distribution with mean $(r+1)/(n+2)$ and variance of the order of n^{-3}, where r is the observed frequency of A in n trials; the result follows by setting $r = n$ and using Chebychev's Inequality). Thus were Popper to take his own prior distributions as hypotheses, he would actually not only find them sometimes falsified, but would also find that in those cases the true priors are those which seem to assign a probability in the neighbourhood of 1 to a hypothesis approaching the strength of a universal generalisation. I think we must conclude that Popper's attempted disproof via (1) of the possibility of probabilistic-inductive confirmation of laws is vitiated.

But Popper's own case against importing synthetic ***a priorism*** into methodology in the form of pseudo-objective priors (or pseudo-logical measure functions) remains, in my opinion, quite sound. It seems to me, as it has to others, no defence of such procedures that under quite general conditions the dependence of the form of the posterior distribution on that of the prior is asymptotically almost everywhere nil: this is the phenomenon Savage calls the principle of stable estimation (Savage *et al.* [1963], though to appreciate the full generality of the result see Blackwell and Dubins [1962]). However, what happens in the limit and what happens after a given time are two different things, and in general the form of the prior will be reflected in that of the posterior distribution.

Where does this leave the philosophy of induction? There are, of course, non-Bayesian theories of inference from data: Popper himself has one (though he would not put it like that); and there are non-Bayesian, non-Carnapian probabilistic theories. I think that these other theories face insuperable difficulties. I am not going to argue this here, but instead simply state that I believe the subjective Bayesian approach quite capable of furnishing an adequate, and I also believe correct, account of inductive inference. I shall list some of its achievements later. First, however, I shall expound that theory in a form free of the defects I referred to in my introduction.

5 SUBJECTIVE PROBABILITY DISTRIBUTIONS

The theory of subjective probability is nowadays usually called Personalism (after L. J. Savage, who coined the term). It is sometimes regarded as a theory about the degrees of belief of some Ideal Reasoner, ideal in the sense first that he can recognise all relationships of deductive entailment, and so is able always to satisfy the conditions that where $\boldsymbol{a} \vdash \boldsymbol{b}$ then $P(\boldsymbol{a}) \leqslant P(\boldsymbol{b})$ (or if $A \subseteq B$ then $P(A) \leqslant P(B)$ in set theoretic language), and that where $\boldsymbol{a} \vdash \boldsymbol{b}$ and $P(\boldsymbol{a}) > 0$ then $P(\boldsymbol{b}/\boldsymbol{a}) = 1$. This faculty immediately endows him with resources which in principle exceed those of any computer, since it is well known that if a finitely axiomatisable subtheory of Peano arithmetic is consistent, which is generally taken for granted, then the decision problem for first order consequence is recursively unsolvable (work is, however, currently going on to investigate weakening the condition of fidelity to entailment relations, and in extending the domain of the probability function to include hypotheses about deductive relationships themselves; see for example Niiniluoto [1983]). He is ideal also in his beliefs being coherent, in de Finetti's sense of satisfying a necessary (but not sufficient) condition for not being vulnerable to inevitable loss relative to some combination of (negative or positive) stakes.

To call the P-value degrees of belief of an ideal reasoner is to imply that they are not the degrees of belief of any actual person. Indeed, it seems very doubtful that subjective probabilities are in any sense psychological entities at all. Admittedly, one is sometimes almost certain that something or other will happen, or quite certain, or not at all certain, and it is a consistent hypothesis that any ordering can be regarded as induced by a real valued function; and this is true for an ordering of future events in terms of subjective feelings of their respective likelihoods (I am not using this word here in its technical meaning). Now while I think that it is certainly true that people do make use of a graduated scale on which in appropriate circumstances they will locate with great precision the likelihoods of certain events, and give or take odds accordingly, I also think there are strong reasons for supposing that this scale does not always measure their belief. Consider the example of two lotteries, one with 10^9 and one with 10^9+1 tickets. Other things being equal people will regard 10^{-9} and $(10^9+1)^{-1}$ as the respective chances of any given ticket winning, and define their odds accordingly. But I do not think that their degree of belief that any given ticket will win in one lottery can be differentiated from their degree of belief that a given ticket will win in the other. Indeed, by making the lotteries sufficiently large we can be sure that our degree of belief that a particular ticket will win is indistinguishable from a conviction that it will not win. An earlier generation of thinkers introduced the term 'moral certainty' to denote a state of belief which cannot differentiate the members of an interval of probability values sufficiently close to 1 or 0 and thereby implicitly to point up the difference in such cases between belief and probability. The conclusion I wish to draw from this discussion can be put succinctly thus: probability (in the sense of epistemic probability) is not of the genus *belief*, but a quantifiable modality of truth whose attributions are simply believed or not believed. The subjectivist position is characterised merely by the refusal to acknowledge any objective foundations for these beliefs.

Two objections might be made to this attempt to distinguish subjective probabilities from grades of belief. One is that the lottery example does not show that an Ideal Reasoner, as opposed to an ordinarily human one, is not capable of refining his degrees of belief to correspond to the probability values. This may be true, but now 'degree of belief of an Ideal Reasoner' is a purely theoretical term, introduced merely to correspond to a probability; and in that case we can dispense with it without loss and simply retain the primary notion of probability. The point I am trying to make here is that there is nothing in many of the characteristic ways we evaluate subjective probabilities that can be represented as psychological measurement. Degrees of belief are invoked simply because no objective ground can be supplied for most if not all these evaluations, and because if indeed they are subjective, then it seems required that some subjective denotation be supplied for them, which can only be some psychological propensity of the agent. But nothing of any theoretical substance is added by endowing subjective probabilities with referents called 'degrees of belief': de Finetti's [1937], the source of most if not all the fundamental results about subjective probabilities, never once finds it necessary to identify them with degrees of belief. Moreover, paradoxes arise from making such an identification, when considering states of null information, for this would seem to imply that the appropriate probability distribution is uniform. But in that case it should be uniform relative to any redescription of the sample space; but it is well known that this is not possible in general where the sample space is a continuum. Indeed, this observation, familiar though it is, seems to me to be quite decisive against the identification of subjective probabilities with belief-states.

The second objection I can conceive being made is that in the lottery example subjective probability is being confused with statistical probability, and that while there is a difference in the physical probabilities with respect to the two lotteries there is no difference in the subjective probabilities of any given tickets taken from each lottery being the winning tickets. But this objection cannot be sustained. All subjective probabilists set subjective probabilities equal to the objective physical probabilities where these are known (I am excepting the extreme subjectivists like de Finetti, who do not believe in physical probabilities, but what I say certainly goes for the remainder) and there is no other relevant information. This is indeed the criterion by which the likelihood term $P(\boldsymbol{e}/\boldsymbol{h})$ is evaluated in computing the posterior probability of $\boldsymbol{h}$, where $\boldsymbol{h}$ is some statistical hypothesis and $\boldsymbol{e}$ sample evidence. It is called the Principal Principle by Lewis [1980], was declared inconsistent by Miller [1966], and had its formal reputation restored by Howson and Oddie [1979].

So we seem to be left with irreducibly subjective prior distributions, but distributions which, if my argument above is correct, should nevertheless not be identified with states of anybody's belief. From the point of view of a subjective Bayesian, these prior distributions function merely as exogeneously determined parameters within an inductive logic constituted by the calculus of probabilities. The subjective Bayesian theory has been extensively criticised, however, precisely because it does not lay down any criteria for the form of prior distribution to be adopted in any given case. The correct answer to this objection is in my opinion indicated by pointing out an interesting analogy between the subjective Bayesian theory and modern deductive logic. Not since the time of Bacon has there been any serious claim that deductive logic should not only provide canons of valid inference, but in addition should lay down criteria on the acceptability of the premisses themselves. Only the former duty is now recognised as legitimate. Similarly, one might argue, it is not obviously the province of a theory of inductive inference to lay down what the initial probabilities (*i.e.*, the prior probabilities) should or should not be; what such a theory should do, on the other hand, is to provide the canons according to which one should proceed from these initial distributions to a final (posterior) distribution.

While this, I think, disposes of or at least defuses the objection, it raises, in its talk of canonical procedures, the 64,000 dollar question. What is canonical about the probability calculus? Why, in other words should we obey its principles? There is a well known argument for obedience to the canons of deductive logic: the price of disobedience is the risk of inferring a falsehood from a truth. No such sanction obtains in the inductive case. Let us first note that, impressive as the argument sounds for the deductive case, it does not actually show, and cannot, that in any given nondeductively valid inference you will actually infer a falsehood from a truth. All it says is that the risk is there. But it does not and cannot quantify this risk, and you may in fact make nothing but invalid inferences and come to no harm. You may even acquire a great deal of knowledge you would not have acquired had you been more scrupulous (after all, science is based on types of inference which are not deductively valid). Knowing that an inference is deductively valid merely supplies a guarantee, though none may in fact be needed, if Nature is sufficiently indulgent. What guarantees, however weak, are afforded by obedience to the probability calculus?

One guarantee is that you will never be led to declare, for example, a hypothesis and its negation both more (or less) probably true than not; you will never declare that $\boldsymbol{a}$ is more probable than $\boldsymbol{b}$ even though $\boldsymbol{a}$ deductively entails $\boldsymbol{b}$, or that $\boldsymbol{a}$ and $\boldsymbol{b}$ are logically equivalent but possess different probabilities, or that the probability of $\boldsymbol{a}$ given $\boldsymbol{a}$ is not equal to that of a logical certainty. Admittedly, certain properties of the calculus are quite conventional, like the choice of the unit interval as the scale relative to which probability is calibrated: any other interval would do as well, theoretically speaking. But there is a residuum which is not conventional. Cox showed in his [1961] that the calculus of probabilities can be obtained simply from the assumptions that $P(\boldsymbol{a} \wedge \boldsymbol{b})$ is some function $\boldsymbol{f}$ of $P(\boldsymbol{a}/\boldsymbol{b})$ and $P(\boldsymbol{b})$, and that $P(\neg \boldsymbol{a}/\boldsymbol{b})$ is a function $\boldsymbol{g}$ of $P(\boldsymbol{a}/\boldsymbol{b})$, where $\boldsymbol{f}$ and $\boldsymbol{g}$ are suitably well-behaved. We need to have a reason, however, for regarding these principles as authoritative. It is not enough to say that they are analytic of the notion of reasonable belief, for this is simply to beg the question.

de Finetti, in a famous argument, shows that if a set of betting quotients over an algebra of hypotheses is immune to a Dutch Book then that set satisfies the (finitely additive) probability calculus. This implies that were you to bet indifferently on or against a prediction at odds $\boldsymbol{p}/1-\boldsymbol{p})$ where $\boldsymbol{p}$ is the probability you attribute to that prediction, and if you were to do this for all the predictions in some set closed under the truth-functional operations, then you would be vulnerable to a Dutch Book if the $\boldsymbol{p}$-values are not the values of a probability function. But it is not immediately obvious that there is any reason why you should bet in this way, or even bet at all. Of course, people can usually be induced to bet if they are given sufficiently favourable odds, but even then the odds that would induce them to bet on and the odds that would induce them to bet against will usually be different. Given this fact one strategy is simply to give up the idea of point-valued probabilities in favour of so-called upper and lower probabilities determined respectively by the maximum of the odds you would give against and the minimum you would give on the events in some field (see for example Smith [1861]). The resulting theory, obtained by the usual Dutch Book considerations, is elegant but weak compared with the theory of point-probabilities (it replaces equations by inequations in general, except where the upper and lower probabilities coincide). I think, however, that it is possible to salvage the original Dutch Book argument by means of the following considerations.

As a first step, let us note that your preferences for some actions over others, each associated with a set of penalties and rewards waiting upon the various contingencies of Nature, would seem naturally to depend jointly on those penalties and rewards and the probabilities you attribute to those contingencies. A bet is simply an action in conditions of uncertainty where the contingencies are restricted often to simple alternatives and the penalties and rewards easily quantified. Now there is a well known argument due to von Neumann and Morgenstern that subject to very plausible conditions your preferences for certain actions over others in the context of a well-defined probability distribution over the various relevant contingencies can be represented on the real line by the values of a set of associated expected utilities. Were the possible actions simply the pair 'bet at odds R : S on *a*', 'do not bet at all', and were the utilities associated with **a**, $\neg$**a** in the former act both in the same proportion to S and R respectively then the expected utility of betting is positive if $R/S < \boldsymbol{p}/(1-\boldsymbol{p})$ where $\boldsymbol{p}$ is the probability you attribute to **a**. In other words one could infer a preference for betting over not betting from a knowledge of S, R and p. But even were one able to derive preferences from utilities in this theory, which it is well known one cannot, the utilities existing only conditionally on a set of antecedently given preferences, this would still be no help. For the von Neumann–Morgenstern theory assumes that the probabilities involved satisfy the probability calculus (the theorem of total probability is involved in deriving the expected utility principle). We on the other hand want to justify obedience to the calculus in terms of rationality constraints placed on the way people give and take odds, and to do this we need independent grounds for establishing a relationship between these odds and their probability assignments.

There are, of course, extant arguments, like that of Savage (Savage [1956]), which develop a theory of subjective probabilities obedient to the probability calculus from consideration of suitably constrained preferences between, essentially, bets. However, Savage's argument is actually a representation-theorem showing that under certain conditions the derivative probability ordering among events is preserved under a mapping of the events into the real line by a (unique) probability function. But I am supposing that the probability numbers are already given, and that they furnish the grounds for the preference ordering. The problem then is to show that if they do not obey the probability axioms, then undesirable consequences may ensue.

It is in fact possible to show that a preference for betting over not betting is determined by what looks, formally speaking, like the expected values of the respective actions, computed from a knowledge of R, S and the individual's probability assignments. To accomplish this we shall assume that you are going to evaluate the desirability of an act which gives you S if **a** is true and $-R$ if not simply on the values of S, R and the probability $\boldsymbol{p}$ you assign to **a**. Let us in fact assume that there exists a real-valued function $U(\boldsymbol{p},S,R)$ by which it is possible to measure the desirability of this act. Let us assume also that this function satisfies the conditions of being linear in each argument (at least for small to moderate values of S and R), that $U(0,S,R)$ is independent of S, that $U(1,S,R)$ is independent of R, that $U(\boldsymbol{p},S,-S)$ is independent of $\boldsymbol{p}$, and that $\partial U/\partial R < 0$ for $\boldsymbol{p} < 1$. It is easy then to show that

$$U(\boldsymbol{p},S,R) = \boldsymbol{c}(\boldsymbol{p}S-(1-\boldsymbol{p})R)+\boldsymbol{d} \qquad (2)$$

for arbitrary constants $\boldsymbol{c},\boldsymbol{d}$ where $\boldsymbol{c} > 0$. It follows that at any price less than $\boldsymbol{p}S$ (where $S > 0$) you would think it desirable to buy at that price the chance of getting S if $\boldsymbol{a}$ is true and 0 if not (for deciding not to buy at price $\boldsymbol{x}$ means that $U(\boldsymbol{p},0,0)$ exceeds $U(\boldsymbol{p},S-\boldsymbol{x},\boldsymbol{x})$).

Now (2) has the form of a positive linear transformation of the expected value of a binomial random variable X where $X[\boldsymbol{a}] = S$ and $X[\neg\boldsymbol{a}] = -R$, but we have not yet required $\boldsymbol{p}$ to be a probability in the sense of the probability calculus; nothing in the derivation of (2) depends on this being assumed (I do not count calibrating probabilities in the closed unit interval with 0 and 1 assigned to the two extremes of certainty as a substantive principle of the calculus: any other interval would do as well, theoretically speaking, granted appropriate modifications of the addition and multiplication laws). It could be objected that in making U a linear function only of $\boldsymbol{p}$, S and R it is implicitly assumed that $\boldsymbol{p}'$, where $\boldsymbol{p}'$ is the probability assigned to $\neg\boldsymbol{a}$, is itself a linear function of $\boldsymbol{p}$, and that the boundary conditions $\boldsymbol{p}' = 0$ (resp. 1) when $\boldsymbol{p} = 1$ (resp. 0) determine the relationship $\boldsymbol{p}+\boldsymbol{p}' = 1$ (in fact we can note that this relationship follows from setting $U(\boldsymbol{p},S,R) = U(\boldsymbol{p}',-S,-R)$). The reply to this objection is twofold. In the first place, that $\boldsymbol{p}$ and $\boldsymbol{p}'$ sum to 1 is weaker than the ordinary additivity condition; and secondly, the multiplication principle is certainly not implicit in (2). An apparently more serious objection is that we have already conceded that odds given on and the reciprocal of the odds against are quite likely to be separated by a nondegenerate interval, yet it follows simply from (2) that they are not. The answer is partly that the assumption that behaviour is determined simply by $\boldsymbol{p}$, S and R in the way indicated is a strong one (people simply are not *homines economici*); more importantly, however, we have assumed that the probabilities themselves are point-valued. In practice, of course, they would not be determined outside an interval. Are we not then left after all with a theory of upper and lower probabilities? Yes and no. Yes where the $\boldsymbol{p}$-values are so separated, no where they are not. In case this sounds too glib, it can be pointed out that in many cases the separation is likely to be sufficiently small to be ignorable, and in particular where background information is such as to make estimates relatively 'objective', as in determining the likelihoods in standard applications of Bayes's Theorem (and as Smith *op. cit.* p. 176 notes, where there are distinct upper and lower prior probabilities, the upper and lower posterior probabilities are actually determined by Bayes-type equations as opposed to inequations).

Granted this, we can infer from (2) that such people will, if they wish to be protected from the possibility of losing money come what may, assign both absolute and conditional probabilities in accordance with the probability calculus (the analogue of (2) for conditional probabilities is obtained by repeating the argument for conditional bets). For it follows from (2) that to bet on $\boldsymbol{a}$ at odds $R:S$ is more desirable than not to bet at all if $U(\boldsymbol{p},0,0) < U(\boldsymbol{p},\boldsymbol{k}S,\boldsymbol{k}R)$ for all positive $\boldsymbol{k}$ (up to that value at which U ceases to give accurate results) *i.e.*, if $R/S < p/(1-p)$, and hence that $p/(1-p)$ is the supremum of the odds you will regard as meriting a bet on **a**, and the infimum of the odds you would regard as meriting taking the other side of such a bet (represented of course by changing the signs of S and R). It is now a straightforward matter to use de Finetti's argument to show that if your probability assignments over an algebra of hypotheses do not satisfy the (absolute and conditional) probability calculus then you may be made to lose money whatever happens. There is an important qualification to be made to this account, however. If the hypotheses in question number among them any which are only one way decidable then the rational odds to declare are either 0 or infinite, and certainly need bear no relation to your probability assignments over those hypotheses.

I shall end with a brief consideration of the question, if the theory of conditional probabilities does after all furnish the logic of inductive inference (together with subsidiary principles which evaluate the likelihood terms in Bayes's Theorem, *etc.*), how responsive are people, and in particular those most concerned with making valid inductive inferences, to its precepts? An affirmative answer, by the way, should not be taken to imply that people consciously employ the calculus or that they consciously evaluate probabilities, at least in any precise way. We are growing accustomed to think that first order logic, or some extension of it, supplies the standard of deductive argument, though such a recognition implies nothing about the way any given person will behave, or even think, when confronted with a question of the validity of an inference. The thesis is rather that the canons of deductive logic operate as constraints on our reasoning activity, but constraints which we can disobey and even be ignorant of, and I believe the same, *mutatis mutandis*, to be true of the calculus of probabilities.

Evidence that these constraints are frequently effective, however, and particularly in science, is furnished by noting that all the standard methodological principles regarding the evaluation of hypotheses in the light of data seem to be explicable as examples of probabilistic reasoning. For example, that empirical evidence is more supportive of a hypothesis which predicts it the less explicable that evidence is by any other plausible alternative hypothesis, would seem to be pretty standard methodology, and it has a simple probabilistic reconstruction (glossing support now as the difference between posterior and prior probabilities). That a hypothesis whose parameters are calculated from the data is less supported by those data, other things being equal, than one which predicts the data directly has also a straightforward probabilistic explanation (see Howson [1984b]). Data from a variety of experimental tests are reckoned to be more informative than evidence obtained by repeating a test of the same type: again, there are straightforward probabilistic explanations of this (see Horwich [1982], p. 118, Urbach [1981], Howson and Franklin [1984]). Finally (though the list could be extended), the peculiar property of statistical hypotheses of not predicting in a deductive way any empirically observable effects renders a Bayesian treatment practically mandatory. The attempts made by various people starting with Fisher to construct a non Bayesian account of statistical inference have admittedly become the ostensible practice of many statisticians, but those theories are certainly anomaly-ridden or else quite arbitrary, and in some cases impossible to put into effect without a good deal of artificiality (references are difficult to supply here in any completeness, spread out as they are over an enormous literature, mostly in the journals, but Hacking [1965] contains a very good critical discussion of Fisher's and the Neyman–Pearson theory). By contrast the Bayesian theory is very natural. I am referring here to the personalist as opposed to the so-called objective Bayesian theory; the latter is altogether a more doubtful affair, with its unnormalisable prior distributions, its problems with lack of invariance under reparametrisation, and its generally indefensible insistence on a unique prior in any given case (for a sharply critical discussion see Seidenfeld [1979], Milne [1984]).

London School of Economics

APPENDIX 'CLASSICAL' PROBABILITY MEASURES

Such a measure is a type of finitely additive probability measure P whose domain is a set $\mathscr{F}$ of subsets of some set S, where $\mathscr{F}$ contains S and is closed under finite unions and complements (and hence under intersections). The members of S will be regarded as the class of models—defined say up to elementary equivalence—of some language L (augmented possibly by some additional conditions), and $\mathscr{F}$ will correspond to the assertions formulable within L (hence the insistence on closure under the Boolean operations). $\mathscr{F}$ will sometimes but not always be the set of all subsets of S. If S is finite it generally will and each point of S will correspond to a strongest assertion or complete theory in L.

If P is a classical measure then S is either finite or there is a 'natural' mapping of S into a bounded region of ε^n, n dimensional Euclidean space. In the former case $P(A)$, where $A \in \mathscr{F}$, is equal to the cardinal number of A divided by that of S. In the latter, $P(A)$ is equal to the length, *i.e.*, Lebesgue measure of the image of A in ε^n divided by that of S. Any bicontinuous mapping from ε^n into ε^n will yield alternative representations of A and S and unless the map is linear the value of P will not be preserved: hence the necessity of selecting essentially one representation of S in ε^n.

REFERENCES

BLACKWELL, D. and DUBINS, L. [1962]: 'Merging of Opinions with Increasing Information', *Annals of Mathematical Statistics*, **33**, pp. 882–6.
COX, R. T. [1961]: *The Algebra of Probable Inference*. Baltimore: Johns Hopkins.
FAGIN, R. [1976]: 'Probabilities on finite models', *Journal of Symbolic Logic*, **41**, 50–8.

DE FINETTI, B. [1937]: 'Foresight: its logical laws, its subjective sources', *Studies in Subjective Probability*, H. E. Kyburg and H. E. Smokler (*eds.*), pp. 97–158.

HACKING, I. [1965]: *The Logic of Statistical Inference*, Cambridge.

HORWICH, P. [1982]: *Probability and Evidence*. Cambridge.

HOWSON, C. [1973]: 'Must the logical probability of laws be zero?', *British Journal for the Philosophy of Science*, June 1973, pp. 153–63.

HOWSON, C. [1984a]: 'Popper's solution of the problem of induction', *Philosophical Quarterly*, April 1984, pp. 143–7.

HOWSON, C. [1984b]: 'Bayesianism and Support by Novel Facts', *British Journal for the Philosophy of Science*, **35**, pp. 254–61.

HOWSON, C. and FRANKLIN, F. [1984]: 'Why do Scientists prefer to vary their Experiments?', *Studies in History and Philosophy of Science*, **15**, pp. 51–62.

HOWSON, C. and ODDIE, G. [1979]: 'Miller's so-called Paradox of Information', *British Journal for the Philosophy of Science*, **30**, pp. 253–61.

JEFFREYS, H. [1948]: *Theory of Probability*. Cambridge.

JEFFREYS, H. and WRINCH, D. [1921]: 'On Certain Fundamental Principles of Scientific Enquiry', *Philosophical Magazine*, **42**, pp. 269–98.

KOLMOGOROV, A. N. [1933]: *Foundations of the Theory of Probability*. New York.

KUIPERS, T. [1978]: *Studies in Inductive Probability and Logic* . Reidel.

LEWIS, DAVID [1980]: 'A Subjectivist's Guide to Objective Chance', *Studies in Inductive Logic and Probability*, **11**, *ed.* Jeffrey, pp. 263–93.

MILLER, D. [1966]: 'A Paradox of Information', *British Journal for the Philosophy of Science*, **17**, pp. 59–61.

MILLER, D. [1974]: 'Popper's Qualitative Theory of Verisimilitude', *British Journal for the Philosophy of Science*, **25**, pp. 166–77.

MILNE, P. M. [1984]: 'A Note on Scale Invariance', *British Journal for the Philosophy of Science*, **34**, pp. 49–55.

NIINILUOTO, I. [1976]: ' On a *k*-dimensional system of inductive logic', in the *Proceedings* of the 1976 Biennial Meeting of the Philosophy of Science Association, *eds.* F. Suppe and P. D. Asquith.

NIINILUOTO, I. [1983]: 'Novel Facts and Bayesianism', *British Journal for the Philosophy of Science*, **34**, pp. 375–9.

POPPER, K. R. [1959]: *The Logic of Scientific Discovery*. London: Hutchinson.

POPPER, K. R. [1972]: *Objective Knowledge*. Oxford.

SAVAGE, L. J. *et al.* [1963]: 'Bayesian Statistical Inference for Psychological Research', *Psychological Review*, **70**, pp. 193–241 (co-authored with Edwards, E. and Lindman, H.).

SAVAGE, L. J. [1956]: *The Foundations of Statistics*. Dover.

SEIDENFELD, T. [1979]: 'Why I am not an Objective Bayesian', *Theory and Decision*, **11**, pp. 413–40.

SHIMONY, A. [1970]: 'Scientific Inference', in *The Nature and Function of Scientific Theories*, *ed.* Colodny, pp. 79–173.

SMITH, C. A. B. [1961]: 'Consistency in Statistical Inference and Decision', *Journal of the Royal Statistical Society*, Series B, 23, pp. 1–25.

URBACH, P. [1981]: 'On the Utility of repeating the "same" Experiment', *Australasian Journal of Philosophy*, **59**, pp. 151–62.

WATKINS, J. [1985]: *Science and Scepticism*. Princeton.

Brit. J. Phil. Sci. 38 (1987), 225–242 *Printed in Great Britain*

Biological Species: Natural Kinds, Individuals, or What?

by MICHAEL RUSE

What are biological species? Aristotelians and Lockeans agree that they are natural kinds; but, evolutionary theory shows that neither traditional philosophical approach is truly adequate. Recently, Michael Ghiselin and David Hull have argued that species are individuals. This claim is shown to be against the spirit of much modern biology. It is concluded that species are natural kinds of a sort, and that any 'objectivity' they possess comes from their being at the focus of a consilience of inductions.

1 *Biologists on Species*
2 *Natural Kinds*
3 *But Are Biological Species Really Natural Kinds?*
4 *Species as Individuals*
5 *Why Species are not Individuals: Biological Objections*
6 *Why Species are not Individuals: Conceptual Objections*
7 *But what then are Species?*
8 *Consequences*
9 *Conclusion*

The status of biological species continues to attract attention and controversy. (See, for instance, Eldredge and Cracraft [1980]; Gould [1979]; Grant [1981a]; Levin [1979]; Mayr [1982]; Wiley [1978, 1980]; Splitter [1982]; Mishler and Donoghue [1982]; Holsinger [1984]; Kitcher [1984]; Eldredge [1985].) There is a strong feeling among biologists, at least there is a strong feeling among most zoologists and somewhat less of one among botanists (a difference to be discussed later), that species are somehow different from the other groupings of organisms we find (or make) in nature. Species, like *Drosophila melanogaster* or *Canis lupus*, are thought to be 'natural', in some way objective or existing independently of the classifier. In this, species differ from the groups (taxa) found at other ranks, for instance that of the genus. The classifier's own thoughts and aims have a much greater role to play in the delimiting of members of these other groups.

But wherein lies the naturalness of species? With the coming of evolutionary theory, traditional answers seem less than adequate. Recently, in a brilliantly innovative move, the biologist Michael Ghiselin [1966, 1969, 1974a,b, 1981; 1987], supported by the philosopher David Hull [1975, 1976, 1978, 1979, 1981], has argued that evolution shows us to have mis-

Received 24 April 1986

construed the nature of species. They are not groups or classes of organisms, like hockey players on a team. Rather, they are integrated *individuals*, with organisms having the relationship to their species of part to whole rather than member to class. And thus properly seen, argue Ghiselin and Hull, the specialness or naturalness of species is self-evident. Species are natural or real in the way any biological individual is natural or real.

In this essay, I argue that, stimulating though the species-as-individuals (s–a–i) thesis may be, it runs counter to much accepted biological thinking, as well as to logic. We must rely on more traditional conceptual tools to establish the naturalness of species. But, with some exceptions and qualifications, this can be done, and the very exceptions and qualifications themselves establish the correctness of the overall approach.

1 BIOLOGISTS ON SPECIES

Let us start with what biologists have to say about species. The most interesting fact is that the category of species can be and is characterised in so many different ways, with corresponding ways of defining particular taxa names. Here, four major species concepts will suffice. (See Mayr [1982] and Grant [1981a] for recent discussions of the multiplicity of species concepts.)

First, we have the most obvious and intuitive concept of all. We find the organic world broken up into groups of similar looking organisms, with gaps between the groups. The concept thus refers to some notion of overall similarity of appearance possessed by organisms within species taxa. In Charles Darwin's words, a species is 'a set of individuals closely resembling each other' (Darwin [1859], p. 52).

Physical nature or *morphology* is the key to this species concept, and it is therefore invoked when one deals with particular species taxa. To be a member of *Homo sapiens* you must be relatively hairless, capable of walking upright, with a large brain, and so forth. In fact, today it is recognised that invariably there is diversity, even within such morphologically delimited groups, so biologists frequently use polytypic or polythetic definitions: lists of features, a combination of which is sufficient for species membership, but no one of which is necessary (Beckner [1959]; Simpson [1961]; Hull [1965]).

Next, we have a concept which has, perhaps a little strongly, appropriated unto itself the title of *biological* species concept. This refers to breeding, or the lack of it. One well-known formulation, due to Ernst Mayr [1942], states that species are 'groups of actually or potentially interbreeding natural populations which are reproductively isolated from other such groups'. You do not normally find taxa name definitions using this concept, but I do not see in principle why not. Specify some individual, say Brigham Young, as your reference point, and then members of the same taxon are potential or actual interbreeders, with some obvious qualifications to take account of sex, and so forth. (See Mayr [1982] for his attempts to give a more refined version of the concept.)

Third, we have a concept which deliberately refers to *evolution*. In the words of the paleontologist G. G. Simpson [1961]: 'An evolutionary species is a lineage (an ancestral-descendant sequence of populations) evolving separately from others and with its own unitary evolutionary role and tendencies.' A taxon name would get a related definition. If we suppose that humans first appeared about 1/2 million years ago, *Homo sapiens* is the name for the group which has descended from the original organisms. (This certainly seems to be the kind of definition that paleoanthropologists have in mind. See Johanson and Edey [1981]; Johanson and White [1979].)

Fourth and finally, we have a concept which does for the world of genes, what the first concept did for morphology. The category concept refers to overall *genetic* similarity clusterings, such clusterings, being separated from others by gaps. Mayr [1969] again: 'When an evolutionary taxonomist speaks of the relationship of various taxa, he is quite right in thinking in terms of genetic similarity, rather than in terms of genealogy.' A particular species name would be defined in terms of genes held in common (together perhaps with information about chromosomes, structure and so forth). For obvious reasons, you do not often see definitions of this ilk, but they do exist. With increasingly sophisticated methods of analysing genomes, their similarities and differences, we might expect to see more such definitions.

There are other concepts which could be and sometimes are invoked, for instance concepts based on ecology (Van Valen [1976]). But, we have enough for our purposes. The question to be asked now is why a taxon which falls into a category characterised in one of the above ways should be thought natural or real in some sense. Since this is a question *about* science, rather than *within* it, we turn to philosophy for guidance.

2 NATURAL KINDS

Traditionally, philosophers have treated the status of species as being part and parcel of a larger question about the reality of 'natural kinds'. Why do we think the whole physical world to be divided into different sorts of things: gold, water, stars, as well as *Homo sapiens* and *Drosophila melanogaster*?

Roughly speaking, there have been two main answers to this question (Ayers [1981]).[1] Credit for the first is given to *Aristotle*. He argued that the world—at least, the world of scientific inquiry—is made up of substances. Any particular substance, like a sample of gold or an individual man, results from the interaction between the substance's underlying matter and its form. This latter gives a substance its nature or essence. Objects of the same kind, like two men, are the same because they have the same form, which is embedded in different samples of matter. Substances have their form essentially, that is to say, one cannot be a substance of a particular kind without having the required form.

---

[1] In my view, most of the modern supporters of natural kinds end up somewhere to the right of Aristotle (*e.g.* Kripke [1972]; Putnam [1975]; Wiggins [1980]). Frankly, I am not sure how far these modern thinkers really intend their ideas to apply to biology, since they generally do not bother to refer to the works of practising taxonomists, and at times show an almost proud ignorance of the organic world. Any comments I have to make against Aristotle apply equally against them. Dupré [1981] shows how ignorant most modern philosophical thinkers are about biological reality.

Crucial to the Aristotelian position is the distinction between a 'real definition' and a 'nominal definition'. The former enables you to define a natural kind name, including the name of an organic species, in terms of attributes which stem necessarily from the very essence of a substance (Aristotle called these attributes 'properties'). Thus, in the case of *Homo sapiens* the essence involves the notion of rationality. Unpacking, we get such properties as the power of speech. A real definition would consequently refer to this power. However, not all attributes of an individual stem from the essence. There are features which are possessed 'accidentally'. Although these features are non-essential, it might nevertheless be possible to distinguish a natural kind using only accidents. In the case of humans, both bipediality and featherlessness are accidents, and yet it so happens that the set of featherless bipeds is one and the same as the set of rational animals. Any characterisation in terms of accidents yields a 'nominal definition'.

The great rival to Aristotle's analysis came in the seventeenth century, from the pen of *John Locke*. He argued that reality lies in the underlying particles which go to make up any particular substance. Locke himself was, in fact, doubtful that we could ever truly know these basic units. But, any real definition would have to make reference to these building blocks, specifically to their shape, structure, motion, composition, and so forth. Surface definitions are simply marks of the structure beneath. Consequently, any surface definition could never be more than nominal. For Locke, a definition of humans in terms of rationality has no more and no less status than a definition in terms of bipediality.

But what of the underlying real structure? Even here Locke wanted to deny the absoluteness of Aristotelian essences. Shapes and so forth can change, taking a substance from one kind to another. Not to mention borderline cases:

> There are Animals so near of kind both to Birds and Beasts, that they are in the Middle between both ... There are some Brutes, that seem to have as much Knowledge and Reason, as some that are called Men ... and so on till we come to the lowest and the most inorganical parts of Matter, we shall find everywhere, that the several Species are linked together, and differ but in almost insensible degrees (Locke [1975], III, vi, 12).

Hence, ultimately for Locke, even definition in terms of reality involves a conscious decision to divide. A Lockean definition is therefore never more than what an Aristotelian would label 'nominal'. Any difference between men and changelings 'is only known to us, by their agreement, or disagreement with the complex *idea* that the name *Man* stands for' (Locke [1975], III, vi, 39).

In short, whereas for Aristotle natural kinds are ontological entities, for Locke they are at best epistemological concepts. You have the objective approach, *versus* the subjective approach. The approach which *finds* natural kinds, and the approach which *makes* them.

3 BUT ARE BIOLOGICAL SPECIES REALLY NATURAL KINDS?

Let us now try to put biology and philosophy together. Does either Aristotle or Locke capture the biologist's sense that species are real or natural?

The simple answer is that, as they stand, neither does. Take Aristotle. He would argue that species are real, because they are natural kinds. But, how then could one get a real definition? The morphological approach to taxa will not do, because if there is one thing that modern biology teaches, it is that evolution promotes morphological diversity. Species members are not all the same. Hence the need for polytypic definitions. But, for something to be an Aristotelian property (as opposed to an accident), it must be possessed by every member of the kind, and distinguishing the group from others. Polytypic definitions are not enough. Hence, morphology will not do. (Mayr [1963]; Dobzhansky [1970]; Dobzhansky *et al.* [1977], discuss variation within species.)

The same considerations apply to genetic features. Evolution promotes genetic diversity (Lewontin [1974]). And, similarly, the features relied on by other approaches to species fail the Aristotelian. Is the sterile worker ant even a potential interbreeder, and would one really want to say that an entirely artificially produced fruit-fly could never really be in *Drosophila melanogaster*?

In any case, after Darwin, strict Aristotelianism simply will not work (Hull [1965]; Mayr [1969, 1982]). Evolution says that you can take virtually any property you like, and if you go back (or forwards) enough in time then ancestors (descendants) did not (will not) have it. But, this is just what Aristotle cannot handle. The whole point about a natural kind is that its properties exist in perpetuity, like mathematical objects. And clearly such properties have to be passed on by (to) ancestors (descendants). Evolution denies this.

What about Locke? Initially, things seem very much more promising. Morphological criteria give you nominal definitions of species names. Alternatively, if you favour phylogenies, you use relationships of descent—something which Locke himself, incidentally, was not that enamoured with. ('[M]ust I go to the Indies to see the Sire and Dam of the one, and the Plant from which the seed was gather'd, that produced the other, to know whether this be a Tiger or that Tea', Locke [1975], III, vi, 23.) Then, genetic criteria give you the closest things you can get to real definitions.

And, contrary to Locke's own doubts, we can know quite a bit about these. Nary an issue of *Science* or *Nature* appears without fresh details of the genetic structure of some organism.

One often finds versions of a neo-Lockean proposal in the literature (for instance, Mayr [1963, 1969]). Unfortunately, you purchase your solution to the species problem at too high a price. You have to relinquish claims to the ultimate objectivity or reality of species. A Lockean natural kind is essentially subjective or arbitrary. And that is just what you do not want to concede, when it comes to species. In some sense, species are real!

We are caught in a dilemma. Evolution refutes Aristotelianism; but, Lockeanism is inadequate. According to Ghiselin and Hull, and an increasingly large number of supporters, we must go back to biology. When we do this, we see that Aristotle and Locke share a false premise. Reject this premise, and hope rises for a solution to the species problem. (Sympathisers with Ghiselin and Hull include Mayr [1976]; Wiley [1978]; Rosenberg [1980]; Sober [1980]; Beatty [1982]; Splitter [1982]; Eldredge and Cracraft [1980]. Intimations of the Ghiselin/Hull approach are to be found in Theodosius Dobzhansky's classic, *Genetics and the Origin of Species*.)

4 SPECIES AS INDIVIDUALS

Aristotle and Locke agree that species are natural kinds. The taxon *Homo sapiens* is a class, with individual humans like Michael Ruse and Charles Darwin as members. I qualify for membership in the class *Homo sapiens*, because I possess certain attributes, whatever they may be. So does Charles Darwin. My dog Spencer does not have these properties, and thus does not qualify. He has his own species, *Canis familiaris*.

Ghiselin and Hull argue that species are not natural kinds at all: They are not classes with members. Rather, species are *individuals*, just as particular organisms are individuals. Hence, just as the relationship between my arm and myself is one of part to whole, rather than member to class, so my relationship to the species *Homo sapiens* is one of part to whole. I, and Charles Darwin, are parts of the human species, just as Spencer is part of the species *Canis familiaris*.

The reformers argue that, once we see species in the true light, all of the problems about species start to fade. Of course, species are real. No one doubts the reality of Michael Ruse, or of Spencer. They are individuals. Real things. Part of the furniture of our world. So are species.

What kind of claim is the s–a–i thesis? It is not solely or even primarily an empirical claim. 'Look! There's an individual!' Rather, it is more of a conceptual claim, whose plausibility must be argued for. Consider a chessboard. You can think of this as an individual, made up of 64 parts, or as a class of 64 squares. It depends on your perspective as to which makes more sense—are you making chess-boards, or are you teaching someone the rules of chess. *The crux of the s–a–i thesis, therefore, is whether modern evolutionary* *biology inclines one to treat species more as individuals, or more as classes, as natural kinds.*

There is of course the initial question as to what precisely one means by an 'individual'. Ghiselin and Hull point to the fact that, whatever else an individual may or may not be, we recognise organisms as paradigm examples of individuals. Organisms are not just diffuse, artificially created entities. They are integrated beings. They have internal organisation. Hence, if we can show that, in important respects, species are like organisms, we can reasonably say that species are individuals.

But, claim Ghiselin and Hull, from an evolutionary perspective species have the very marks of individuality that organisms have. Just like organisms, species come into being, exist for a period in space and time, and then go. And, they have at least some sense of organisation. As the leading evolutionist Ernst Mayr said (before he himself was converted to the s–a–i thesis): '*Species are the real units of evolution*, they are the entities which specialize, which become adapted, or which shift their adaptation' (Hull [1976], p. 183, quoting Mayr [1969]).

That evolutionary theory treats species as individuals becomes clear when we look at their uniqueness. Adolf Hitler was an individual—as such (unlike his diaries) he came uniquely and went uniquely. A copy of Adolf Hitler is not he. It cannot be. Similarly, we have this uniqueness for species. 'If a species evolved which was identical to a species of extinct pterodactyl save origin, it would still be a new, distinct species' (Hull [1978], p. 349). If you are a species, you simply cannot be born again, any more than Adolf Hitler can be.

What about change? Organisms can undergo major change, and still be the same organism. What counts is continuity. The limits of an organism are birth and death. The same is true of species. 'There is no limit to the genetic change that can take place in a species or population before it becomes extinct or speciates' (Hull [1976], p. 182). In fact, just like an organism, so long as the continuity persists, we have the same species.

Finally, let us mention one revealing point. Biologists take one specimen from a species, using it as the marker. The species name (which, as with all individuals is a proper name) is attached to this marker—the type specimen—by an act which is akin to baptism. This specimen does not have to be a 'typical' member of the species, whatever that might mean. And, of course it does not. The type is part of the whole, not a member in the class.

> The fact that any specimen, no matter how atypical, can function as the type-specimen makes no sense on the class interpretation; it makes admirably good sense if species are interpreted as individuals (Hull [1976], 175).

All in all, whatever 'common sense' may say, modern evolutionary biology demands that species be regarded as individuals. Hence, the naturalness of species.

5 WHY SPECIES ARE NOT INDIVIDUALS: BIOLOGICAL OBJECTIONS

Ingenious though it is, the Ghiselin/Hull attempt to slice through the Gordian knot constraining the species problem fails. There are several significant reasons why species cannot properly be considered as individuals.[1]

First, look at matters at the most basic biological level. We think organisms are individuals because the parts are all joined together. Charles Darwin's head was joined to Charles Darwin's trunk. But, in the case of species, this is not so. Charles Darwin was never linked up to Thomas Henry Huxley. Of course, you might object that although Darwin's head was never linked directly to his feet, they were linked indirectly through intermediate parts. Analogously, as evolutionists presumably we believe that Darwin and Huxley were linked by actual physical entities (namely, the succession of humans back to their shared ancestors). But, this objection fails, for the point is that these links have now been broken and lost. If (gruesome thought!) Darwin's head were physically severed from his feet, we would certainly have no biological individual.

Yet, with justice, Ghiselin and Hull will respond that these speculations are beside the point. The required condition for individuality is not mere spatio-temporal contiguity. It is rather some sort of internal integration or organisation. Because of such internal organisation, the U.S. is one country, even though Alaska and Hawaii do not touch any of the other states. The fifty states work together, in a way that (say) the forty-eight mainland states together with Ontario and Quebec do not. Analogously, Charles Darwin is an individual, not because of spatio-temporal contiguity, but because his parts are organised, working *together*. The same is true of other biological entities, even those which break into parts at some points in their life cycles, like slime molds. *And*, the same is true of species. They have an integrating organisation, with the parts contributing to the whole.[2]

But, this will not do, at least not in the light of much modern thought about the working of evolution. First and foremost, thinking of a species as an integrated individual goes flatly in the face of the way in which the major evolutionary mechanism of natural selection is generally regarded today. Selection leads to adaptations, features which help organisms in life's struggles for survival and reproduction. But who precisely benefits from adaptations? Is it the possessors alone, or do others benefit? In short, at what level of biological organisation does selection work? Is it between individuals, benefiting individuals, or is it between higher entities like species, benefiting species taken as a whole.

---

[1] Other critics of the s–a–i thesis include Caplan [1980, 1981] and Kitts and Kitts [1979]. Unfortunately, these critics revert to a modern-day, genetic, Aristotelian essentialism. I find myself agreeing with much in Hull's [1981] spirited response to them.

[2] Could spatio-temporal contiguity alone count as the criterion of individuality? We surely think of the planet Earth as an individual on these grounds. But, while this may be true, we do not think of Earth as a *biological* individual, which notion is the focus of the s–a–i thesis. Incidentally, however, given plate-tectonic theory and the consequent claims about Earth's organisation, a case might be made for Earth's geological individuality, transcending mere spatio-temporal contiguity.

Until recently, most people casually assumed that selection could work at virtually every level of biological organisation. In particular, one could have selection between groups of organisms, including between species. The units of selection, in vital respects, were species. As Mayr (1969) said: 'Species are ... the entities ... which become adapted.' However, majority opinion today is that selection just does not work in this way. As Charles Darwin himself argued, ardently, selection works chiefly if not exclusively at least at the level of the individual organism. 'Group selection' at the level of the species does not work. A species is not adapted. An organism (or, at most, a limited number of organisms) is. Any species effects are just epiphenomena on individual effects, or at most, on population effects. (See Brandon and Burian [1984], for a review of this topic, and Ruse [1980] for Darwin's views on the subject.)

If this is all so, then there is something very odd indeed about speaking of a species as an individual. It is very far from being an integrated unit like an organism. The individual organisms of a species are all working for their own benefits, against those of others. Any species cooperation, any species integration, is secondary on the particular organism's self-interests. And, in any case, one is hardly likely to get species-wide secondary effects. Cooperation will, at most, be between relatives, or fellow population members. Generally, selection pits organisms against each other (although not necessarily in a crude 'nature red in tooth and claw' fashion).[1]

Individual selection and the s–a–i thesis simply do not go together. What about obvious counters? Some biologists believe that group selection can work. This is true, but hardly makes the s–a–i-thesis again compelling. Group selection supporters think it works for populations, not species, and no one denies the importance of individual selection. (See Wilson 1975 and Wade 1978 for recent views on group selection.) Conversely, some biologists argue that the true 'individual' in individual selection is the gene, not the whole organism (Dawkins 1976). Does not my argument prove too much, suggesting that organisms should not be considered true individuals—which is clearly absurd? But, while this point does show that for some biologists the level of individuality does not necessarily stop at the whole organism level, no one denies that organisms (thanks to selection) are sufficiently well organised to be considered individuals in their own right. Richard Dawkins [1982], for instance, speaks of organisms as 'vehicles' which carry within them the units of selection, 'replicators' or genes. An organism, to such a biologist, is no less an individual than a BMW is to a racing driver.

---

[1] One does get cooperation between organisms. But, the point is that ultimately, biology regards it as 'enlightened self-interest'. Hence, at root we have tensions—separate reproductive strategies—between organisms. At times, for instance where mates are involved, these tensions break right out. See Trivers [1971]; Wilson [1975]; Barash [1977]; Clutten Brock [1982]; and Ruse [1979b], for more details.

Continuing with biological counters to the above critique, what about Steven Stanley's (1979) notion of species selection, where it is suggested that trends are a function of the success or failure of species? Again there is little help for the s–a–i thesis. Even if one accepts species selection, and many would not, the key operation of natural selection is with the individual. Drawing attention to the trends that one often sees in the fossil record, Stanley suggests that there is nevertheless a randomness about the members of new species with respect to a trend. Although a trend may (say) be from smaller to bigger, a new species in the line of descent could well have small-bodied members. (No doubt, if one persisted, one could devise some form of species selection where the group as a whole was significant. But its realisation in nature is obviously another matter. See Arnold and Fristrup [1982].)

Perhaps the strongest biological case for the s–a–i thesis comes through the notion of a species as a number of organisms sharing a common 'gene pool', with shared types of genes being passed on to common ancestors (*i.e.* a kind of hybrid notion formed from several of the species concepts. Dobzhansky [1970]; Dobzhansky *et al.* [1977]). Here you might think we have the kind of integration required for individuality. Certain genes flow between the organisms of a species, and between no others. But, this hardly denies the key importance of individual selection. Moreover, there are today strong questions about the biological importance, at the species level, of such genetic sharing. It was once thought that gene flow, between populations, is a key factor in keeping the organisms of a species alike. Now, it seems more likely that normalising selection is the key causal factor. Species members sit on the top of the same 'adaptive peak'—if they vary too much from the species norm, then selection wipes them out.

This downgrading of the significance of gene flow is a most important point, because (being itself one which comes from modern evolutionary theorising) it strikes right at the heart of the claim that the s–a–i thesis (however counter-intuitive it may seem) must be accepted on the basis of modern biology. John Endler's [1977] already-classic study brings both theoretical and empirical evidence to bear demonstrating the restricted effects of gene flow. Basically, gene flow would be expected due to migration ('the relatively long-distance movements made by large numbers of individuals in approximately the same direction at approximately the same direction at approximately the same time', p. 182) and dispersal ('the roughly random and nondirectional small-scale movements made by individuals rather than groups, continuously, rather than periodically, as a result of their daily activities', p. 181). One would expect that migration would be a most effective way of uniting a species, even if widely dispersed; but, as Endler points out, this is rarely so, since migration is usuall accompanied by return migration and organisms give birth in the place where they were themselves born. Such 'philopatry' has obvious adaptive virtues—birds may migrate to winter feeding grounds but they return to already-established and proven breeding grounds.

Dispersal is *a priori* a less promising way in which gene flow might be greatly effective, and there are a number of reasons why its importance should not be over-estimated: ethological, ecological, and physiological inadequacy of hybrids between distant species members; random loss of new gene forms because they are rare; infrequency of long-distance travellers (as opposed to migrators); and more (pp. 28–9). All in all, therefore, one should not over-emphasise the unity of the species because of the supposed circulation of shared genes. (Similar points are made by Grant [1981a,b]; Levin [1981]; and Mishler and Donoghue [1982]. Although the point just made applies to both animals and plants, since the latter are spatially more fixed, expectedly the 'genetic integration' argument has always seemed less plausible to botanists.)

So, once again we come back to the individual organism and to its response to the environment (including fellow species members). If you take Darwinian selection seriously, you simply must reject the s–a–i thesis. Note that I am certainly not denying that the members of a species are frequently 'united' in having similar causal pressures, whether these be genetic or selective or whatever. Of course they do. That is what makes them part of the same species. The question is whether this 'unification' is significantly more than similar causes. The s–a–i-supporter has to say that it is, even to the point of the kind of integration we find in individual organisms, and this is what I deny. (Caplan [1980], rightly emphasised that same causes lead to same effects, and this accounts for species members being similar.)

6 WHY SPECIES ARE NOT INDIVIDUALS: CONCEPTUAL OBJECTIONS

Let us move on to more conceptual-type objections to the s–a–i thesis. Crucial to the thesis is the claim that, logically, a species can appear only once. If it dies, that is it. In this, species are just like paradigm individuals organisms. Adolf Hitler cannot be resurrected. Neither can extinct species of pterodactyl.

I will leave Christians to fight their own battles about human bodily resurrection. As far as species are concerned, time and technology have shown the s–a–i claim wrong. Today, through recombinant DNA techniques and the like, biologists are rushing to make new life forms. Significantly, for commercial reasons the scientists and their sponsors are busily applying for patents protecting the new creations. Were the origins of organisms things which uniquely separate and distinguish them, such protections would hardly be necessary. Old life form and new life form would necessarily be distinct. Since apparently they are not, this suggests that origins do not have the status claimed by the s–a–i boosters. (See Wade [1979, 1980a,b].)

Relatedly against the s–a–i thesis, in crucial respects it seems that it really does not treat species and their organisms all that very differently from the old way of treating of species as classes, with members included according to the possession of certain required properties. Take an organism. How do you know that my hand is part of the individual, Michael Ruse? Because it is joined on—that's why! But my dog Spencer certainly is not joined on, in the same way, to the species *Canis familiaris*. So why do we want to say that he is part of the species? Because he descended from the original ancestors, along with the rest of the group—that's why!

Descent is starting to look very much like an essential property. Spencer is part of the group *Canis familiaris*. Indeed, we even seem to have a real/nominal distinction at work here. In Spencer's case, if challenged about his status, I can in fact produce papers attesting to parentage. But in the case of the other four-legged being that lives in my house, I have no such documentary evidence about origins. And yet, I am as sure that Sesame is a cat, as that Spencer is a dog. Why? Because she looks and behaves like one. She miaows, purrs, keeps alloof, jumps from heights, stays up half the night, and is fastidiously clean. In short, she has all the identifying marks of cattiness (*Felis domestica*).

Clearly what I, and everyone else, am doing here is employing morphological and related criteria. Of course, Ghiselin and Hull recognise and appreciate the use of such criteria. They simply refuse to give such use any significant theoretical status (see Ghiselin [1981]). What can this all mean, but that one is using nominal criteria, because real essence's descent relationships are unknown? Hence, for all the talk, the s–a–i thesis treats species as classes, with descent giving real essence and with morphology giving nominal definitions.

There are other objections you can raise against the s–a–i thesis. One is that it has controversial implications about the temporal limits of a species, and the possible evolutionary change that such a unit can encompass. So long as one gets no new group breaking off from an evolving lineage, one has one and only one species, whatever the change (just as one has one and only one organism, despite the change from caterpillar to butterfly). The cladistic school of taxonomy would accept this implication, but most biologists would not. They distinguish, for instance, between *Homo habilis*, *H., erectus*, and *H. sapiens*, despite the lack of branching. *Homo habilis* had a brain size of around 700 cc, much closer to that of a gorilla (*Gorilla gorilla*) at 500 cc, than to modern man at 1400 cc. Hence, species divisions are made—divisions which cladists and s–a–i supporters must ignore. (Cladism is discussed well in Eldredge and Cracraft [1980], and Wiley [1981].)

Indeed, the s–a–i thesis is more extreme than cladism. Cladists end a species as soon as there is any branching within the group. But, if just a small population broke from a parent species, leaving the parent unaffected, one would have a situation very similar to an asexual organism budding off a small part. The parent remains. Similarly, the s–a–i thesis would have to count an analogous parent species the same original individual. Carried to the extreme, classification could become very difficult indeed.[1]

And, finally, let me point to an implication which has rather drastic consequences for the social sciences. Laws of nature generally do not refer to specific spatio-temporally bounded objects. They are rather 'timeless regularities in nature'. But, if species are individuals, then any claims about the organism of species, restricted to the species, cannot be laws. In particular, any claims exclusively about human beings cannot possibly be laws. Hence, at one stroke, the social sciences, as they stand today, cease to be sciences in any worthwhile sense. To say the least, this is a somewhat drastic consequence.

The conclusion is clear. There is no absolute reason to treat species as individuals, and compelling reasons not to do so.

7 BUT WHAT THEN ARE SPECIES?

Either species must be groups, or they must be individuals. There is no third option. They are not individuals, hence they must be groups. But, we have seen that species cannot have the absoluteness of Aristotelian natural kinds. Evolution makes this impossible. The question which therefore remains is whether we can raise species above the rather subjective level of Lockean kinds? Are species more than just artificial collections of organisms?

They are indeed. Moreover, the reasons why we rightly think that species are more than artificial collections—why we think species are natural—are similar to reasons why we think there are natural groups encountered elsewhere in science, for instance in chemistry and geology.

The key reason why species are properly treated as natural kinds lies in that most distinctive fact noted earlier: the multiplicity of species concepts and of possible definitions of taxa names. To see the connection between naturalness and multiplicity, let us pull back for a moment and ask a general question about science. At what point is it in science that we feel we are onto something 'real'? When is it we accept that we are not just dealing with hypothetical figments of a creative scientist's imagination? The strong consensus is that the breakthrough comes when we put *together* two or more *different* areas of theory into one unified whole. If you have two different subjects, and they are joined, so that the one complements the other, and *vice-versa*, then you are inclined to think that there's more than mere chance at work. Somehow, the unified theory tells you about the real. Such a coming together, could not be mere coincidence—especially, if you can spell out the unification in terms of some overall theory (Leplin, [1984]).

---

[1] An escape would be to embrace the neo-saltationary theory of 'punctuated equilibrism', supposing that one gets periods of stasis, followed by abrupt switches from one species to another. See Eldredge and Gould [1972]. The fears just expressed vanish. Wake [1980] and Mishler and Donoghue [1982] note just how tied up the s–a–i thesis is with this theory, and significantly the s–a–i thesis has been embraced enthusiastically by Eldredge [1985]. But, there are serious queries about the position. See Gingerich [1976, 1977]; Maynard Smith [1981]; and Ruse [1982] for an overview.

This unification, known philosophically as a 'consilience of inductions', was the one thing that Darwin always mentioned, when his theory was challenged. 'I must freely confess, the difficulties and objections are terrific; but I cannot believe that a false theory would explain, as it seems to me it does explain, so many classes of facts' (Darwin [1887], 1, 455). And, it remains important today. For instance, in the recent geological revolution, people accepted plate tectonics when they saw that different areas of geology are unified in the one theory (Ruse [1981]).

What about classification? Is there a possibility of some sort of consilience here, separating the natural or real from the merely arbitrary? William Whewell, who had the distinction of being both a professor of mineralogy and of moral philosophy, thought there was. A natural classification is one where different methods yield the same results. Particularly, if you have reasons for the coincidence, you feel sure that the classification cannot be just chance.

> The Maxim by which all Systems professing to be natural must be tested is this:—that the ***arrangement obtained from one set of characters coincides with the arrangement obtained from another set.*** (Whewell [1840], 1, 521, his italics. For more details on how Whewell used his ideas in mineralogy, see Ruse 1976.)

And, modern philosophers agree with Whewell. Thus Hempel:

> The rational core of the distinction between natural and artificial classifications is suggested by the consideration that in so-called natural classifications the determining characteristics are associated, universally or in a high percentage of all cases, with other characteristics, of which they are logically independent. (Hempel [1952], 53. See also Schlesinger [1963] for more references to this criterion.)

Coming back to organic species, we see that we have a paradigm for a natural classification. There are different ways of breaking organisms into groups, and they *coincide*! The genetic species is the morphological species is the reproductively isolated species is the group with common ancestors. Moreover, there are reasons for the coincidence. As the zoologist Mayr points out, bringing several of the definitions together:

> The reproductive isolation of a biological species, the protection of its collective gene-pool against pollution by genes from other species, results in a discontinuity not only of the genotype of the species, but also of its morphology and other aspects of the phenotype produced by this genotype. This is the fact on which taxonomic practice is based (Mayr [1969], 28).

Note, moreover, that the coincidence between variously delimited species is not unexplained. Certain genes do lead to certain morphological effects, and so forth. The consilience fits within overall biological thinking.

This consilience then is the reason why it is reasonable to think of species as natural kinds. Like the natural kinds of other sciences, they demand our attention, not because they represent some ultimate essentialist ontological carving up of the real world, but because they unite different criteria of division. They may not be Aristotelian kinds. But they are more than Lockean kinds. (See also Ruse 1973.)

8 CONSEQUENCES

A number of questions arise. What about the real/nominal definition distinction? It vanishes—which is a good thing, because it is an outmoded Aristotelian holdover anyway. You might argue that genetic differences are more crucial than anything else. (Caplan [1980, 1981]; and Kitts and Kitts [1980] argue just this.) But, from an evolutionary perspective, the genes do not have this kind of privileged status. If organisms do not have the right morphology, they will fail, no matter how superior their genotype. And in any case, a consilience is like a quarrel or a tango. You must have at least two parties. Hence, it really does not make much biological or philosophical sense to say that genes are more essential than morphology. Or that any other single feature is the 'true' essence of a species.

Do we still have laws about species members? I do not see why not. The solution I am offering affirms the existence of natural kinds, albeit not Aristotelian kinds. This means that it is still open to everyone to make universal claims about the members of particular species. And, these claims can rise above the merely contingent or happenstance. This does not mean that every claim that has been made about human beings embodies genuine laws. I am not, for instance, defending the validity of every part of Freudian psychoanalytic theory. On the other hand, such theory is not being ruled out as a genuine science, on *a priori* grounds, before we even start.

Finally, let me make brief reference to some of the ongoing concerns that biologists have about species. I must emphasise that I am not trying to offer a quick and easy solution to every biological species query. Species which were difficult to evaluate before this essay, will be as difficult to evalute after this essay. I am trying to show why biologists, generally speaking, think that species are natural. But, the obverse side of the coin is that when difficulties arise and biologists no longer feel anything like as convinced of the naturalness of certain groups, the analysis offered above should show why. In particular, doubts about the reality of species should arise when the various ways of defining species names come apart, and fail to coincide.

This is indeed the case. As mentioned earlier, botanists often find themselves less than convinced of the reality of many plant species. Why? Simply, because so often plant groups which morphologically and ecologically and in other ways seem to be good species, fail the test of reproductive isolation or some like thing. One just does not have the consilience required for naturalness. Conversely, when there is isolation, there are sometimes few other differences. In such cases, and in analogous cases in the animal world, species—however drawn—are not considered that natural. It is interesting to note how, in the case of so-called 'sibling species', where members of different reproductively isolated groups are morphologically similar, morphological differences are eagerly sought. Much relief is felt when such differences are found. (Grant 1981a has an excellent discussion on the difficulties plants raise for species concepts. Mayr [1963], Dobzhansky [1970], and Dobzhansky *et al.* [1977], discuss in full detail difficulties arising in the animal world.)

Also, the analysis I have offered shows just why it is that biologists are far less convinced of the reality of taxa of higher levels, than they are of species taxa. There simply is not the required consilience. Reproductive barriers are irrelevant, and there are no measures of morphological difference to coincide with genetic difference, to coincide with evolutionary difference. If anything, the evidence is that such measures are impossible. Hence, although 'species are made by God, higher taxa are made by man'.

9 CONCLUSION

Ghiselin and Hull are surely right when they argue that we must break with Aristotelian essentialism in biology. After Darwin, such a position is otiose. But they go too far when they deny that biological species are natural kinds of any sort. They are such kinds, and the reason why it is reasonable for us to accept them as such is that same reason which makes us think any scientific claim goes beyond the merely hypothetical. It is because species are consilient.

ACKNOWLEDGEMENTS

A version of this paper was given at a conference on Natural Kinds, sponsored by Simon Fraser University, February 1983. The following friends have read and criticised (sometimes extensively criticised) that version: Michael Ayers, Art Caplan, Michael Ghiselin, Vernon Grant, Marjorie Grene, David Hull, James Lennox, Elisabeth Lloyd, Ernst Mayr, Alex Rosenberg, Laurence Splitter, Paul Thompson. I am grateful also to an anonymous referee of this journal for constructive comments.

University of Guelph, Ontario, Canada

REFERENCES

ARNOLD, A. J. and FRISTRUP, K. [1982]: The theory of evolution by natural selection: a hierarchical expansion. *Paleobiology*, **8**, pp. 113–29.

AYERS, M. R. [1981]: Locke versus Aristotle on natural kinds. *Journal of Philosophy*, **78**, pp. 247–72.

BEATTY, J. [1982]: Classes and cladists. *Systematic Zoology*, **31**, pp. 25–34.

BECKNER, M. [1959]: *The Biological Way of Thought*. New York: Columbia University Press.

BRANDON, R. and BURIAN, R. [1984]: *Genes, Organisms, Populations*. Cambridge, Mass.: MIT Press.

CAPLAN, A. [1980]: Have species become declassé? In P. Asquith and R. Giere (*eds.*), *PSA 1980*, **1**, pp. 71–82. East Lansing: Philosophy of Science Association.

CAPLAN, A. [1981]: Back to class: a note on the ontology of species. *Philosophy of Science*, **48**, pp. 130–40.

DAWKINS, R. [1976]: *The Selfish Gene*. Oxford: Oxford University Press.

DAWKINS, R. [1982]: Replicators and vehicles. King's College Sociobiology Groups (*eds*), *Current Problems in Sociobiology*, pp. 45–64. Cambridge: Cambridge University Press.

DARWIN, C. [1859]: *On the Origin of Species*. London: John Murray.

DARWIN, F. [1887]: *The Life and Letters of Charles Darwin, Including an Autobiographical Chapter*. London: Murray.

DOBZHANSKY, TH. [1951]: *Genetics and the Origin of Species*, 3rd ed. New York: Columbia University Press.

DOBZHANSKY, TH. [1970]: *Genetics of the Evolutionary Process*. New York: Columbia.

DOBZHANSKY, TH. *et al.* [1977]: *Evolution*. San Francisco, W. H. Freeman.

DUPRÉ, J. [1981]: Natural kinds and biological taxa. *Philosophical Review*, **90**, pp. 66–90.

ELDREDGE, N. [1985]: *The Unfinished Synthesis*. New York: Oxford University Press.

ELDREDGE, N. and GOULD, S. J. [1972]: Punctuated equilibria: an alterntive to phyletic gradualism. In T. J. M. Schopf (*ed.*), *Models in Paleobiology*. San Francisco: Freeman, Cooper.

ENDLER, J. A. [1977]: *Geographic Variation, Speciation, and Clines*. Princeton: Princeton University Press.

GHISELIN, M. [1966]: On psychologism in the logic of taxonomic controversies. *Systematic Zoology*, **15**, pp. 207–15.

GHISELIN, M. [1969]: *The Triumph of the Darwinian Method*. Berkeley: University of California Press.

GHISELIN, M. [1974a]: A radical solution to the species problem. *Systematic Zoology*, **23**, pp. 536–44.

GHISELIN, M. [1974b]: *The Economy of Nature and the Evolution of Sex*. Berkeley: University of California Press.

GHISELIN, M. [1981]: Categories, life, and thinking. *Behavioral and Brain Sciences*, **4**, pp. 269–313.

GHISELIN, M. [1987]: Species concepts, individuality, and objectivity. *Biology and Philosophy*, **2**, pp. 127–44.

GINGERICH, P. D. [1976]: Paleontology and phylogeny: patterns of evolution at the species level in early Tertiary mammals. *American Journal of Science*, **276**, pp. 1–28.

GINGERICH, P. D. [1977]: Patterns of evolution in the mammalian fossil record. In A. Hallam (*ed.*), *Patterns of Evolution, As Illustrated by the Fossil Record*, pp. 469–500. Amsterdam: Elsevier.

GOULD, S. J. [1979]: A quahog is a quahog. *Natural History*, **88**, pp. 18–26.

GRANT, V. [1981a]: *Plant Speciation*, 2nd ed. New York: Columbia University Press.

GRANT, V. [1981b]: The genetic goal of speciation. *Biologisches Zentralblatt*, **100**, pp. 473–82.

HEMPEL, C. G. [1952]: *Fundamentals of Concept Formation in Empirical Science*. Chicago: University of Chicago Press.

HOLSINGER, K. [1984]: The nature of biological species. *Philosophy of Science*, **51**, pp. 293–307.

HULL, D. L. [1965]: The effect of essentialism on taxonomy;: two thousand years of stasis. *British Journal for the Philosophy of Science*, **15**, pp. 314–26, **16**, pp. 1–18.

HULL, D. L. [1975]: Central subjects and historical narratives. *History and Theory*, **14**, pp. 253–74.

HULL, D. L. [1976]: Are species really individuals? *Systematic Zoology*, **25**, pp. 174–91.

HULL, D. L. [1978]: A matter of individuality. *Philosophy of Science*, **45**, pp. 335–60.

HULL, D. L. [1979]: The limits of cladism. *Systematic Zoology*, **28**, pp. 417–38.

HULL, D. L. [1981]: Kitts and Kitts and Caplan on species. *Philosophy of Science*, **48**, pp. 141–52.

JOHANSON, D. and EDEY, M. [1981]: *Lucy: The Beginnings of Humankind*. New York: Simon and Schuster.

JOHANSON, D. C. and WHITE, T. D. [1979]: A systematic assessment of early African hominids. *Science*, **203**, pp. 321–30.

KITCHER, P. [1984]: Species. *Philosophy of Science*, **51**, pp. 308–35.

KITTS, D. B. and KITTS, D. J. [1979]: Biological species as natural kinds. *Philosophy of Science*, **46**, pp. 613–22.

KRIPKE, S. A. [1972]: Naming and necessity. In D. Davidson and G. Harman (*eds.*), *Semantics of Natural Language*, pp. 253–355. Dordrecht: Reidel.

LANDE, R. [1980]: Genetic variation and phenotypic evolution during allopatric speciation. *American Naturalist*, **116**, pp. 463–79.

LEPLIN, J. [1984]: *Scientific Realism*. Berkeley: University of California Press.

LEVIN, D. A. [1979]: The nature of plant species. *Science*, **204**, pp. 381–4.

LEVIN, D. A. [1981]: Dispersal versus gene flow in plants. *Annals of the Missouri Botanical Garden*, **68**, pp. 233–53.

LEWONTIN, R. [1974]: *The Genetic Basis of Evolutionary Change*. New York: Columbia University Press.

LOCKE, J. [1975]: *An Essay Concerning Human Understanding*, P. H. Nidditch (*ed*.). New York: Oxford University Press.

MAYNARD SMITH, J. [1981]: Did Darwin get it right? *London Review of Books*, 3(11), pp. 10–11.

MAYR, E. [1942]: *Systematics and the Origin of Species*. New York: Columbia University Press.

MAYR, E. [1963]: *Animal Species and Evolution*. Cambridge, Mass.: Belknap.

MAYR, E. [1969]: *Principles of Systematic Zoology*. New York: McGraw-Hill.

MAYR, E. [1976]: Is the species a class or an individual? *Systematic Zoology*, **25**, p. 192.

MAYR, E. [1982]: *The Growth of Biological Thought*. Cambridge: Harvard University Press.

MISHLER, B. D. and DONOGHUE, M. J. [1982]: Species concepts: a case for pluralism. *Systematic Zoology*, **31**, pp. 503–11.

PUTNAM, H. [1975]: The meaning of meaning. In H. Putnam, *Mind, Language, and Reality*. Cambridge: Cambridge University Press.

QUINE, W. V. [1969]: Natural kinds, in *Ontological Relativity and Other Essays*. New York: Columbia University Press.

ROSENBERG, A. [1980]: *Sociobiology and the Preemption of Social Science*. Baltimore: Johns Hopkins University Press.

RUSE, M. [1973]: *The Philosophy of Biology*. London: Hutchison.

RUSE, M. [1976]: The scientific methodology of William Whewell. *Centaurus*, **20**, pp. 227–57.

RUSE, M. [1980]: Charles Darwin and group selection. *Annals of Science*, **37**, pp. 615–30.

RUSE, M. [1981]: What kind of revolution occurred in geology? P. Asquith and I. Hacking (*eds*.), *PSA 1978*, **2**, pp. 240–73. East Lansing: PSA.

RUSE, M. [1982]: *Darwinism Defended: A Guide to the Evolution Controversies*. Reading, Mass.: Addison-Wesley.

SCHLESINGER, G. [1963]: *Method in the Physical Sciences*. New York: The Humanities Press.

SIMPSON, G. G. [1961]: *Principles of Animal Taxonomy*. New York: Columbia University Press.

SOBER, E. [1980]: Evolution, population thinking and essentialism. *Philosophy of Science*, **47**, pp. 350–83.

SPLITTER, L. [1982]: *Natural Kinds and Biological Species*. Unpublished Oxford D.Phil. Thesis.

STANLEY, S. M. [1979]: *Macroevolution. Pattern and Process*. San Francisco: W. H. Freeman.

VAN VALEN, L. [1976]: Ecological species, multispecies, and oaks. *Taxon*, **25**, pp. 233–9.

WADE, M. J. [1978]: A critical review of the models of group selection. *Quarterly Review of Biology*, **53**, pp. 101–14.

WADE, N. [1979]: Supreme Court to say if life is patentable. *Science*, **206**, pp. 664.

WADE, N. [1980a]: Supreme Court hears argument on patenting life forms. *Science*, **208**, pp. 31–2.

WADE, N. [1980b]: Court say lab-made life can be patented. *Science*, **208**, pp. 1445.

WAKE, D. B. [1980]: A view of evolution, *Science*, **210**, pp. 1239–40.

WHEWELL, W. [1840]: *The Philosophy of the Inductive Sciences*. London: Parker.

WIGGINS, D. [1980]: *Sameness and Substance*. Oxford: Blackwell.

WILEY, E. [1978]: The evolutionary species concept reconsidered. *Systematic Zoology*, **27**, pp. 17–26.

WILEY, E. [1980]: Is the evolutionary species fiction?—A consideration of classes, individuals, and historical entities. *Systematic Zoology*, **29**, pp. 76–80.

WILEY, E. [1981]: *Phylogenetics: The Theory and Practice of Phylogenetic Systematics*. New York: John Wiley.

WILSON, E. O. [1975]: *Sociobiology: The New Synthesis*. Cambridge, Mass.: Belknap.

Brit. J. Phil. Sci. 38 (1987), 243–257 *Printed in Great Britain*

Review Article

EXPLANATION AND CAUSATION*

1 *Introduction*
2 *Low Probability Explanation*
3 *The Principle of the Common Cause*
4 *Causal Processes and Interactions*
5 *Conclusion*

1 INTRODUCTION

Wesley Salmon's important new book takes up where his [1970] statistical relevance theory of explanation left off. Salmon [1970] rejected both halves of Hempel's [1965] symmetry thesis between explanation and prediction. Salmon argued that a theory can explain an event it says was very improbable (so not every explanation is a potential prediction); he also tried to show why some predictions are not potential explanations. The positive account underwriting these two conclusions was the idea that the explanation of an event must assemble all and only those events that are statistically relevant to it. A rationale offered for this suggestion was that statistical relevance and causal relevance are the same; the statistical relevance theory was an attempt to give substance to the idea that explanation proceeds by uncovering causes.

Salmon subsequently realised that statistical relevance is not the tool he had hoped it would be. Statistical relevance and causal relevance are different. In the present work, Salmon retains a revised version of the statistical relevance idea, though it now is demoted in importance. Salmon's new theory has two levels. The statistical relevance 'basis' sets the stage for a separate and overtly causal analysis of the processes that produced the event in question. Indeed, the demotion may be even more extreme than this. Although statistical relevance is sometimes described as an essential part of the whole, at other times Salmon follows Humphreys [1981] in thinking that facts about statistical relevance may be quite unnecessary (p. 265); at still other times, he suggests that the statistical relevance basis is merely *evidence* for the causal story (p. 192) and this would suggest that it is not an essential part of an ideally complete explanation at all. Whatever the exact relation is between these two levels, it is quite clear that the main innovation separating Salmon [1984] from Salmon [1970] centres on the concept of cause.

---

* Review of Wesley Salmon [1984]: *Scientific Explanation and the Causal Structure of the World*. Princeton University Press. *xiv*+305 pp. Hardback £37.70. Paperback £12.10. ISBN 0–691–07293–0.

The first four chapters contain Salmon's arguments against what Salmon calls the epistemic and modal conceptions of explanation. The former holds that an explanation merely allows one to infer the truth of the *explanandum* statement; the latter contends that an explanation must show why the *explanandum* had to occur. In subsequent chapters, Salmon sets forth his ideas on causal processes and interactions and on what he calls the principle of the common cause. The book also has some very interesting material on scientific realism, on the foundations of probability, as well as a wealth of examples that repay careful consideration. In the following sections, I will limit myself to Salmon's views on low probability explanation, on the principle of the common cause, and on the concepts of causal process and interaction.

2 LOW PROBABILITY EXPLANATION

Salmon advances two sorts of reasons for his claim that an explanation may imply that the explained event had low probability. The first appeals to examples and to our intuitions about what may be plausibly said about them. In one instance, Salmon describes a simple Mendelian genetic system in which a mating between pea plants of given genotypes is expected to produce red and white blossomed offspring in the familiar 3 : 1 ratio. Salmon claims that if the Mendelian account explains why one offspring plant has red blossoms (conferring on this event a probability = 0.75), it must also explain why another comes out white (conferring a probability = 0.25). 'I find it hard to believe that we do not understand the occurrence of white blossoms in this population just as adequately as we understand the occurrence of red,' he concludes (p. 86).

Salmon's second reason has to do with what Kyburg [1961], in a slightly different context, called 'conjunctivitis'. If explanation must confer high probability (which may fall short of unity), then it is possible to explain each of two events without being able to explain their conjunction. The reason is that the conjuncts may have probabilities that exceed the required threshold, whereas the conjunction falls below. 'The moral to be drawn . . .,' Salmon concludes (p. 87), is 'that there is no reasonable or nonarbitrary way of answering the question "How high is high enough?"

Both these considerations would be stopped dead, if we adopted the modal conception, which Salmon ascribes to von Wright [1971], Harré and Madden [1975], and Mellor [1976], that explanations must confer a probability of unity. But this, apparently, would amount to the idea that statistical theories, whether at the micro- or the macro-level, are without explanatory import. With Salmon, I find this judgement unintuitive.

Nevertheless, a small clarification of what the theory of explanation is really about casts doubt on one half of Salmon's conclusion. A philosophical account of explanation might aim at clarifying what an ideally complete scientific explanation is; or it might try to say when a story has some (nonzero) degree of explanatory power. This is the difference between explicating '*C* ideally explains *E*' and '*C* explains *E*', a distinction that Salmon recognizes (p. 261). The literature on explanation has almost always focused on the first idea. Hempel [1965], it will be recalled, insisted that genuine explanations include laws, even though scientists often make explanatory remarks that fail to mention them. These latter Hempel called 'explanation sketches'. Salmon follows Hempel in trying to explicate a scientific ideal.

We now may divide in two the question about low-probability explanation. Do such stories provide *some* degree of explanation? Are they ideally complete explanations? I think the answers are, respectively, *yes* and *no*. The first of these I take to be obvious and it is this affirmative answer, I think, that Salmon has capitalised on in his criticism of Hempel. Mendelian genetics does not leave us utterly in the dark as to why white-blossomed offspring have the colour they do. A genetic mechanism is specified and a probability distribution of outcomes derived; surely this provides *some* illumination of why these plants are white.

But consistent with this is the idea that explanatory power declines as the *explanandum* becomes more and more improbable. Deterministic explanation is the ideal, high probability explanation second best, and low probability worst of all. Salmon insists that a given theory explains the events it says are improbable no worse than it explains the ones it says are probable. But I suspect this idea is the distorted image of a quite different, and quite correct, observation: If I hand you a story and say that it is the true explanation of both *X* and *Y*, then you perhaps will judge that *X* and *Y* are equally well explained, for what could be better than the truth? But this hardly settles the question of how, if at all, probability and explanatory power are connected.

Let us imagine, instead, that you have to choose between two possible explanations of a given event, without knowing in advance which is true. The likelihood of an explanation is the probability it confers on the observations.[1] It is standardly viewed as a measure of the degree to which a hypothesis is supported by, or *explains*, the observation. Likelier explanations are more explanatory (all else being equal, of course). This is why determinism is an explanatory ideal.

Someone might reject the idea that explanatory power is adequately measured by likelihood. I do not think that likelihood is all there is to explanatory power. I do claim that it is a useful measure. Salmon himself has little to say about factors affecting explanatory power. He briefly discusses

---

[1] It is important not to confuse the likelihood of a hypothesis (*H*), relative to an observation (*O*), with the probability of the hypothesis, given the observation; the distinction is between $\Pr(O/H)$ and $\Pr(H/O)$.

Greeno's [1971] information-theoretic measure, notes some problems that it encounters (pp. 88, 99–100) and suggests that it may be correct when supplemented with causal considerations (p. 126). Salmon observes that Greeno's measure is such that 'the increase in information is maximal when all of the conditional probabilities are either zero or one' (p. 99).

Not only does Salmon reject the demand for deterministic explanation and the demand for high probability explanation; he also rejects the idea, defended by Van Fraassen [1980] among others, that an explanation of *E* must explain why *E*, *rather than some contrasting alternative* *C*, occurred and must do so by showing that *E* was more probable than *C*. Salmon grants that in explaining *E*, we must situate that event in a set of contrasting alternatives. The explanation of *E*, he concedes, presupposes that *E* is the only member of this set that is true. Salmon claims that Van Fraassen's 'fundamental error' is to infer from this that in giving an explanation, we have explained why *E* rather than *C* is true (p. 110).

It isn't entirely clear what Salmon's argument is here. Perhaps the point is simply that we can explain an event without showing that it was more probable than the alternatives. This doesn't conflict with the idea that explaining why *E* rather than *C* requires that *E* be shown to be more probable, since one could claim that explaining *E* is a well-defined and different explanatory problem from explaining why *E* rather than *C*, for any value of *C*. Nor does it conflict with the idea that an explanation is more powerful to the degree that it discriminates between the focal event and its contrasting alternatives.

Can determinism be an explanatory ideal if our current best theory of the world suggests that determinism is false? Our explanatory ideals can be frustrated by the world—deterministic theories *are* more explanatory, but we have reason to think that they are not true. Salmon is right that statistical theories are not explanatorily worthless and that even low probability explanations offer some illumination. But he is wrong, I think, in his assertion (which, I emphasize, is not grounded in an account of explanatory power) that high probability explanations are no more illuminating than low, and that explanations that discriminate among alternatives are no more explanatory than ones than do not (Sober [1984b]).

3 THE PRINCIPLE OF THE COMMON CAUSE

During the same period that Salmon believed that causal relevance could be identified with statistical relevance, he also followed Reichenbach [1956] in thinking that two or more correlated events (which do not cause each other) must trace back to a common cause, such that the common cause induces a particular probabilistic relation among the correlates: it must render them conditionally probabilistically independent of each other. In doing this, the common cause is said to 'screen off' each correlate from the rest; the common cause and the joint effects are thereby said to comprise a 'conjunctive fork' (Salmon [1975]). The idea that correlated events which do not cause each other must have a common cause of this sort, Salmon called 'the principle of the common cause'.

Salmon [1978] subsequently was convinced by Van Fraassen [1980] and others that this principle is violated by quantum mechanical anomalies. The Einstein–Podolski–Rosen thought experiment and its successors provide powerful reasons for thinking that some correlated events need not have (and arguably cannot have) a screening-off common cause. Salmon describes this change of heart in the present work, and discusses 'interactive forks' in which common causes do not screen off. Salmon's book takes the position that, were it not for the causal anomalies thrown up by quantum theory, the principle of the common cause would be acceptable.

Salmon does not sufficiently separate ontological from epistemological formulations of this principle. When taken as a claim about the *existence* of causes of a certain sort, the principle is thrown in doubt by quantum anomalies. But suppose we formulate it as the claim that it is always *prima facie reasonable* to treat correlated events that do not cause each other as stemming from a screening off common cause. Quantum anomalies do not challenge this piece of advice; a methodological principle is not refuted by showing that it sometimes leads to falsehoods. For this reason I think the importance of quantum mechanics has been greatly exaggerated, when it comes to the epistemology of common cause explanation (Sober [1984a]). This is not to say that the principle is correct as methodology; as we will see below there are reasons quite independent of Einstein, Podolski, and Rosen for thinking the principle is off the mark.

Salmon claims the principle of the common cause serves both an explanatory and an inferential function. It describes how known facts may be assembled into a satisfying explanation of a correlation; and it implies that the existence of a common cause ought to be postulated to account for a correlation (p. 211). Focusing on this second function first, we should ask *why* common causes ought to be postulated to explain correlations. What makes them superior to separate cause explanations?

Salmon's answer is tht common causes are plausibly invoked because, otherwise, the correlation would simply be too improbable. Suppose you toss a coin repeatedly while I do too and that each of us obtains heads roughly half the time. If my result is almost always the same as yours on each round, reason cries out for a common cause that coordinates the pairs of outcomes. The idea that the repeated matchings are due to chance is simply not plausible (p. 249).

Note that Salmon here evaluates the quality of explanations in terms of how they probabilify an outcome. The common cause explanation is plausible, he says, because it makes the repeated matchings probable. Here we find Salmon favourably disposed to the thought that explanatory power increases with likelihood.

Examples like this, convincing in themselves, may obscure the fact that separate cause explanations can also sometimes explain correlations. Let us switch examples to see how. If two species have some trait in common, one might explain this by appeal to the idea that they inherited their resemblance from a common ancestor, or one might say that the trait evolved independently. The common cause explanation asserts that the matching characters are *homologous*; the separate cause explanation says they are *homoplasious* (the result of parallelism or convergence). Even if 'reason cries out for common causes' in the coin-tossing case, nothing quite so straightforward can be said in this evolutionary context. A common cause scenario would explain the matching of the two species, but so, apparently, would a separate cause account.

Call the two observed species S_1 and S_2. Let A_1 be a postulated ancestor of S_1 and A_2 be a postulated ancestor of S_2. In each case, we assume the descendent species obtained its characteristic from the postulated ancestor. The homology explanation asserts that $A_1 = A_2$; the homoplasy explanation asserts that $A_1 \neq A_2$. Both these explanatory stories can account for the observed correlation. Moreover, the existence of A_1 and A_2 screens off the character states of descendent species from each other, ***regardless of whether the ancestors are identical with each other***. In recognizing interactive forks, Salmon realised that common causes need not screen-off. Here is a case that shows that screening-off is not a virtue unique to common cause explanation.

The contrast between homology and parallelism helps clarify what a common cause explanation actually is. Two species with homologous characteristics trace those traits back to a single (token) ancestor; however, if the traits are parallelisms (like the wings of a bird and the wings of a bat, for example), they trace back to the *same type* of etiology (*e.g.*, selection for flight), even though there is no common ancestor from which both obtained the novelty. Inferring common ancestors is a case of common cause explanation in the sense of common token, not (just) common type. Although most of Salmon's examples involve common token causes, his discussion of molecular theory (pp. 213–27) is not of this sort, a point that Van Fraassen ([1980], p. 123) had already made; Salmon cites this objection but seems not to answer it (p. 226–7).

Although Salmon discusses numerous examples of screening-off common cause explanations, the examples are all in the same vein. A correlation cannot be dismissed as simply due to chance, and a scientific discovery provides the details of the explanatory common cause. Salmon does not discuss cases in which a detailed separate cause explanation is in serious scientific contention. The evolutionary problem of inferring common ancestry for different species on the basis of their similarities and differences is of this sort. It provides a test—I think an unfavourable one—for Salmon's approach.

Above I argued that screening-off is not the unique prerogative of common cause explanation. This means that if correlated events are best

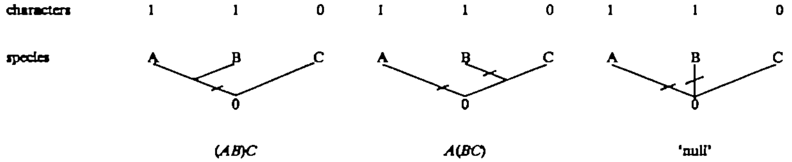

Figure 1 When species *A* and *B* exhibit the derived character state (1) and *C* the ancestral (0) form, the genealogy (*AB*) *C* can account for this observation by postulating a single evolutionary change, whereas *A*(*BC*) and the 'null' genealogy require at least two each.

explained by postulating a common cause, the reason cannot be that only common causes screen-off. I now want to suggest that there is an even more fundamental difficulty facing the principle of the common cause. Not only is screening-off not the key to the virtues of common cause explanation; in addition, correlations do not always cry out for common cause explanation.

Most evolutionists are wary of inferring common ancestry simply because two species resemble each other. Cladists, for example, hold that resemblance with respect to *some* characteristics (*i.e.*, so-called 'apomorphic'—derived—characters) is evidence of common descent, whereas resemblance with respect to *other* ('plesiomorphic'—*i.e.*, ancestral) traits is not.[1] The grounds for this methodological distinction are still a matter of debate. But it is not controversial that this methodology would be appropriate, *if* evolutionary changes were always very improbable.

To see why, suppose species *A*, *B*, and *C* are known to share a common ancestor. The question is whether any two of these species share a common ancestor that neither shares with the third. Two alternative pair-wise groupings are shown as (*AB*)*C* and *A*(*BC*) in Figure 1. The third, 'null', genealogical hypothesis asserts that no two form a group.

Suppose the ancestor common to all three—the 'root' of the tree—has characteristic 0. Suppose further that *A* and *B* are observed to have the derived (apomorphic) trait 1, whereas *C* has the plesiomorphic state 0. We will abbreviate this triplet of observations by saying that the characteristic has a 110 distribution. Which of the three genealogical hypotheses is best supported by this observation?

Cladists assert that the best supported grouping is (*AB*)*C*, since it requires only one evolutionary change (indicated by a slash) to account for the observations. Note that *A*(*BC*) and the null genealogy each require at least two changes. It is for this reason that cladists have used 'parsimony' as the name for their method. The best supported hypothesis is the one that requires the fewest homoplasies.

The most parsimonious explanation of 110 asserts that the matching species—*A* and *B*—have a common ancestor unique to them. It is consistent with this hypothesis that the character states of these two species were

---

[1] See Sober [1983], [1984a] for discussion of the controversy currently underway concerning the phylogenetic inference problem.

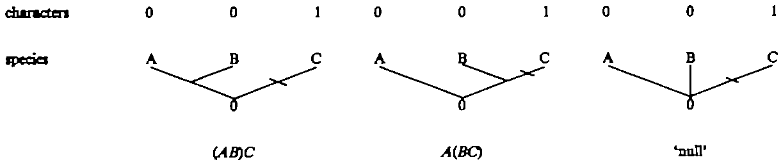

Figure 2 When species *A* and *B* exhibit the ancestral character state (0) and *C* the derived (1) form, the three genealogies can each account for this observation by postulating a single evolutionary change.

obtained as a homology from that ancestor. The alternative hypotheses, on the other hand, assert that *A* and *B* separately derived their apomorphies. It is in this sense that (*AB*)*C* suggests a common cause and *A*(*BC*) and the null hypothesis a separate cause explanation of the character distribution.[1]

Notice that if we had observed 001, instead of 110, parsimony would not conclude that (*AB*)*C* is better supported than the alternatives. Figure 2 shows that each of the genealogical hypotheses can account for this observation by postulating a single evolutionary change. None is more parsimonious than the others.

Cladists believe that 110 is evidence favouring (*AB*)*C* but that 001 is evidence for nothing in particular. What could justify this pair of judgments about evidential meaning? Clearly the following assumption suffices: suppose that evolutionary changes—from 0 to 1 or from 1 to 0—are very improbable. If this is true, then parsimonious genealogies will be most likely; they will strain our credulity least.

It is now a matter of controversy whether this sufficient condition for justifying parsimony is also a necessary one. Some hold that parsimony requires uniformly low rates of evolution; others have argued that parsimony requires no such thing. We need not enter into the details of this further debate to recognise some consequences for Salmon's principle of the common cause.

It is clear that if evolutionary changes are improbable, then mere correlations of characters are not, *per se*, evidence of common ancestry. Suppose we have a data set in which there are a few characters distributed as 110 and a great many distributed as 100. If we attend just to overall correlation (taking account of all traits), it will emerge that *B* and *C* are more correlated than are *A* and *B*. Yet parsimony asserts that the data prefer (*AB*)*C* to *A*(*BC*).

It is not just in evolutionary theory that correlated events do not always cry out for common cause explanation. Any two quantities that have monotonically increased over time will be correlated. The price of bread in England and the sea level in Venice are correlated, but no one thinks this is because they are effects of a common cause. Whether two correlates are better explained by common cause or by separate cause explanations is not settled by the mere fact that they are correlated. Additional background information is essential here.

---

[1] If any two species we might be called upon to classify trace back to a common ancestor, the choice between common cause and separate cause explanation, at the minimum, must involve three species—not whether *A* and *B* have a common ancestor, but whether they have a common ancestor that neither shares with some 'outgroup' species *C*.

One of the lessons of the raven paradox, I think, is that confirmation is a three-termed relation between evidence, hypothesis, and a background theory. A black raven may confirm, disconfirm, or be irrelevant to the hypothesis that all ravens are black, depending on what else one believes about the sampling situation (Good [1967]). In posing the raven paradox, Hempel [1945] adopted the 'methodological fiction' that we face the world with no background theory. Knowing nothing but deductive and inductive logic, we are supposed to be able to say whether black ravens confirm or disconfirm. We now know that there is no such problem for inductive logic to address. Observations do not have evidential meaning, save in the context of a background theory.

A similar diagnosis attaches to Salmon's principle of the common cause. Salmon sets his problem the way Hempel set his: we observe a correlation and ask ourselves whether this confirms a common cause explanation. Salmon formulates a principle that says common cause explanations are always better suported than separate cause explanations. It is not just that the screening-off idea fails to fully justify this principle; the additional point is that the principle is not something we should try to justify in the first place. Correlations do not have evidential meaning, save in the context of a background theory.

If this general point is correct, it would be a mistake to try to replace Salmon's principle with another of comparable generality. The phylogenetic inference problem provides a context in which it may be possible to state *a* principle of the common cause, which may be justified in virtue of broad features of the evolutionary process. But these features will be contingent, and so it is to be expected that other, nonevolutionary, processes will not satisfy them. In those cases, inferences to common causes must be given a different rationale. The upshot will be not one, but several, principles of the common cause, each grounded in specific theoretical assumptions about the nature of the processes under investigation.

Given all this, the puzzlements of the Einstein–Podolski–Rosen paradox are mere icing on the cake. This is not to deny that action-at-a-distance is profoundly unsettling. But the connection of screening-off and common causes is less intimate than Salmon supposes. Even without the quantum theory, there is considerable reason to doubt Salmon's principle of the common cause.

4 CAUSAL PROCESSES AND INTERACTIONS

There are many questions that might be addressed in developing the idea that all explanation is causal explanation. Salmon's interest is in the concept

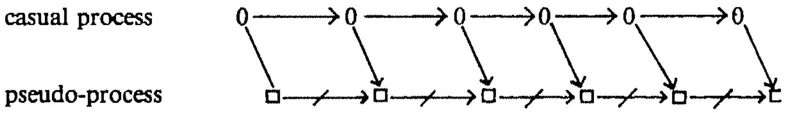

Figure 3 The stages of a pseudo-process are not related to each other as cause to effect, but are each effects of the stages of a different causal process.

of cause itself. However, he decisively reorients this line of inquiry by taking causal processes and their interactions, rather than causing events and their effects, as fundamental. Two questions must be asked about Salmon's proposals: Are the accounts of these causal concepts adequate? How do they illuminate the concept of causal *explanation*? I'll take these up in turn.

A spotlight is rotated in a large domed building, so that a circle of light traverses the ceiling. If the ceiling is high enough and the rotation performed fast enough, the circle will move across the ceiling faster than the speed of light. But special relativity says that causal processes cannot go faster than the speed of light. Hence, the moving circle of light must be a *pseudo*-process. Salmon tries to explicate the difference between a genuinely causal process and a pseudo-process.

Stipulating that a process must conform to relativity theory is not enough. For even if the image on the ceiling moved at subluminal velocities, it still would be a pseudo-process. Another example is supplied by the shadow cast on the road shoulder by a moving car. The moving shadow is a pseudo-process. But why?

Intuitively, the stages of a pseudo-process are not related to each other as cause to effect. Rather, they each are effects of a process at work elsewhere. The structure is that shown in Figure 3. Salmon does not offer this intuitive description, but tries to characterize the difference between causal and pseudo-processes. He takes as primitive the concept of a process (subsuming both causal and pseudo-processes together), of two processes intersecting at a place-time, and modality.

A causal process is one that can 'transmit' a 'mark'. Both quoted terms require illumination and we must be vigilant that neither renders the whole project circular. Here is Salmon's principle of mark transmission:

> Let P be a process that, in the absence of interactions with other processes, would remain uniform with respect to a characteristic Q, which it would manifest consistently over an interval that includes both of the space-time points A and B ($A \neq B$). Then, a mark (consisting of a modification of Q into Q'), which has been introduced into process P by means of a single local interaction at point A, is transmitted to point B if P manifests the modification at B and at all stages of the process between A and B without additional interventions (p. 148).

Does this principle explain why the moving circle of light on the domed ceiling counts as a pseudo-process? Salmon points out that if we place a disk of red glass a foot away from the dome at a point along the beacon's path, then the spot behind the glass will be red. But the redness of this one location is not transmitted to subsequent locations. Nor could other local modifications of the circle of light be transmitted. This, intuitively, is why the image path is a pseudo-process.

The counterfactual formulation of this principle was introduced, Salmon says, to block a counterexample due to Nancy Cartwright. Suppose just before the disk of red glass near the ceiling is introduced, a red filter is placed over the spotlight itself. Then the place behind the disk will become red and subsequent locations of the ceiling image will be red as well. Has one now marked the image path and thereby shown that it is a genuine causal process? Not according to Salmon's principle: 'In this case, the moving spot would have turned red because of the lens even if no marking interaction had occurred locally at the ... [ceiling]' (p. 149).

Notice that Salmon offers us a sufficient condition for mark transmission; he says that a mark is transmitted from point *A* to *B* *if* the process before *A* is in a certin kind of state (*i.e.*, one which will not change, unless there is an intersection with another process) and then changes state at point *A* due to a single interaction, where the new state is present all the way to *B*. Yet, in his application to Cartwright's example, Salmon concludes that his principle shows why no mark is transmitted. Shall we assume, then, that the 'if' was a slip of the pen, a necessary and sufficient condition being the intent of the principle?

The trouble is that the condition is not necessary, though with one clarification, it does seem to be sufficient. The principle applies only to processes on which exactly one causal or pseudo-process impinges (at point *A*), which are before then in states that would persist if no interactions took place. But surely marks can be transmitted when many causal processes are at work and when the process is not initially in the special sort of state that *Q* exemplifies. Salmon's principle is as restricted as a Newtonian principle would be if it were limited to objects moving with constant velocity that then are acted upon by a single component force.

Cartwright's example involves two causal or pseudo-processes: the red glass near the ceiling *and* the filter on the beacon. I do not see how Salmon's solution flows from his principle. However, in his discussion of the example, Salmon uses a quite separate idea to get the desired result. He assumes that causes are necessary for their effects; the ceiling glass is not the cause, Salmon says, because the subsequent red ceiling images would have occurred without it. But cases of overdetermination make this assumption dubious. If Holmes and Watson simultaneously shoot Moriarty through the heart, each of them can be said to have caused his death. Yet neither was necessary.

As to the principle's sufficiency, some constraints on what a mark is are needed. When the glass is placed near the ceiling, it becomes true that all subsequent stages of the image on the roof have the property of occurring after the glass was bolted in place. Moreover, the subsequent stages would not have had this property if the glass had not been affixed. To avoid having to conclude that the image path is a causal process because it has thus been 'marked', the idea of a mark must be clarified further (Kitcher [1985]). Perhaps effects must be restricted to "occurrent" rather than 'historical' properties, in some sense.

I have suggested that Salmon's principle may be true if taken as providing a sufficient condition, but that it is mistaken when understood as supplying a necessary condition. However, circularity seems to threaten whichever reading we consider. The principle describes a process that intersects with just one other process at a given space-time point. *Process*, as I have said, subsumes both causal and pseudo-processes and space-time intersection is taken as a primitive concept. If neither of these concepts is specifically causal, is the charge of circularity misplaced?

Suppose I am interested in the theory of knowledge and introduce a concept called *phy-ing*. An individual is said to phy a proposition if he either knows it or uses the proposition as an ingredient in making a pudding. I then propose that someone knows a proposition if and only if he phies it. It is pointed out that it is impossible to make a pudding out of a proposition, and so the equivalence of knowing and phy-ing is established. I furthermore suggest that phy-ing is not a specifically epistemic concept, since it subsumes both knowledge and something else (food preparation).

Pseudo-processes cannot transmit marks, but causal processes can; intersections with pseudo-processes cannot produce changes, but intersections with causal processes can. I wish to understand these differences in a noncircular way. So I introduce the concept of a process, which subsumes them both, and then say that a causal process is a process which can be changed via a space-time intersection with another process. Process, I observe, is not a specifically causal concept, since it subsumes both causal processes and something else.

I am not certain how closely the epistemology of phy-ing parallels Salmon's theory of causal processes. If we know at the outset that causal processes can do things that pseudo-processes cannot, then a process that does those things must be a causal process. But this does not seem to provide a noncircular account of the causal concept.

Let us waive these worries about the notion of causal process. Suppose we could explain why the moving spot on the ceiling and the car's shadow are pseudo-processes. How would this help in constructing causal explanations? Evidently, we causally explain an event by tracing it back to the cause(s) that produced it. This tracing back must be along the trajectories of causal processes. but which processes should we pursue in giving a causal explanation?

A causal process, recall, is something that *can* be marked. So, if an electric light is connected by wires to many switches, each of which could be used to turn it on, then each of the wirings counts as a causal process, in Salmon's sense. The reason is that each is a channel along which a signal can be transmitted. But only one of the switches is thrown; it is this one, and not the others, that counts as the cause. What are we to find in Salmon's theory that takes us from possible causes to actual ones?

If causal processes are characterised in terms of their ability to transmit marks, then perhaps we should see the marks themselves as causes. A mark occurs when two processes causally interact, where this notion is characterised by a principle similar to the one stated above. the idea is that two processes causally interact at a time if each subsequently exhibits a state different from the one it had before, where the earlier states were ones that would have persisted if no intersection with another process had occurred (p. 171).

Here again Salmon states his principle as a sufficient condition, although he sometimes uses it to explain cases in which no causal interaction obtains. I'll assume in what follows that he meant 'if and only if'. My objections will follow the pattern of those posed before: although a sufficient condition is provided, it is very restrictive, since the condition is not necessary.

Must causes (*i.e.*, causal interactions) produce changes? And must these changes be such that they would not have happened without their causes? I have already asserted that causes need not be necessary for their effects. But I also think that effects do not have to be changes at all.

Consider a Newtonian example: a billiard ball is moving at constant velocity. Is its state at one time a consequence of its earlier state? I do not see why not. Salmon (p. 105) repeats the standard remark that in Newtonian physics constant velocity can have no explanation. This is, I think, a mistake. Supose a billiard ball in uniform motion were bombarded by a multiplicity of influences that cancelled out each other. Many component causes impinged, but the net force was zero. The ball persisted with the velocity it had before. But this subsequent state was not without its causes. After all, forces impinged, and what are component forces, if not component causes? This example also is problematic for Salmon's theory. No mark is transmitted here, because no change in state occurs. But surely we have here the interaction of several causal processes.

Another limitation on the generality of Salmon's principle is that a marked process must retain *the* property that marked it. Introducing a red filter into a beam of light turns the light red, which it then retains, unless marked again. But introducing a gifted teacher into the life of a violin prodigy also may leave its mark, but perhaps not by the retention of any *single* property that is carried forward for the duration of the musical career.

Interpreting Salmon's two principles as providing necessary conditions also conflicts with the idea that causal connections do not require an underlying determinism. Read as necessary, they assert that interactions require changes in state and that causal processes, once marked, must persist in that marked state if no further processes impinge. But in a probabilistic universe, no such requirements can be sustained. A mark may vanish simply because it is stochastic; and interactions at best will imply probability distributions of subsequent possible states, without implying the necessity of any of them.

Several of these objections would dissolve if Salmon's principles were taken at their word—as sufficient conditions, not necessary ones. But then the theory would avoid counterexamples only by succumbing to a radical incompleteness. In addition, Salmon's diagnosis of pseudo-processes and cases of noninteraction would be cut free from his own theory.

I prefer to view Salmon's account as suplying sufficient conditions that cover a relatively simple range of cases. Time will tell whether the theory can be extended fruitfully. Much will depend on the account of counterfactuals that can be developed. The reduction of causal to more general modal concepts remains a tantalising prospect.

5 CONCLUSION

This is a book rich in philosophical ideas. Philosophers interested in causality and explanation in the sciences simply must study it. It is a proper successor to Hempel's theory, whether or not one agrees with its major conclusions. I have mentioned that in Salmon's theory, a process, once marked, is never the same (unless marked again). In the longstanding historical process of philosophical reflection on causation and explanation, Wesley Salmon has undoubtedly left his mark.[1]

ELLIOTT SOBER
University of Wisconsin, Madison

REFERENCES

GOOD, I. [1967]: 'The White Shoe is a Red Herring', *British Journal for the Philosophy of Science*, **17**, p. 322.
GREENO [1971]: 'Explanation and Information.' Reprinted in Salmon [1971], pp. 89–104.
HARRÉ, R. and MADDEN, E. [1975]: *Causal Powers*. Basil Blackwell.
HEMPEL, C. [1945]: 'Studies in the Logic of Confirmation', Reprinted in Hempel [1965b], pp. 3–47.
HEMPEL, C. [1965a]: 'Aspects of Scientific Explanation', in Hempel [1965b], pp. 331–496.
HEMPEL, C. [1965b]: *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. Free Press.
HUMPHREYS, P. [1981]: 'Aleatory Explanations', *Synthese*, **48**, pp. 225–32.
KITCHER, P. [1985]: 'Two Approaches to Explanation', *Journal of Philosophy*, forthcoming.
KYBURG, H. [1961]: *Probability and the Logic of Rational Belief*. Wesleyan University Press.
MELLOR, H. [1976]: 'Probable Explanation', *Australasian Journal of Philosophy* **54**, pp. 231–41.
REICHENBACH, H. [1956]: *The Direction of Time*. University of California Press.
SALMON, W. [1970]: 'Statistical Explanation', reprinted in Salmon [1971).
SALMON, W. (*ed.*) [1971]: *Explanation and Statistical Relevance*. University of Pittsburgh Press.

---

[1] My thanks to the National Science foundation for financial support and to Ellery Eells, Philip Kitcher, and Wesley Salmon for comments on an earlier draft.

SALMON, W. [1975]: 'Theoretical Explanation', in S. KORNER (*ed.*), *Explanation*, pp. 118–45. Blackwell.

SALMON, W. [1978]: 'Why Ask "Why?"' *Proceedings and Addresses of the American philosophical Association*, **51**, pp. 683–705.

SOBER, E. [1983]: 'Parsimony in Systematics: Philosophical Issues', *Annual Review of Ecology and Systematics*, **14**, pp. 335–58.

SOBER, E. [1984a]: 'Common Cause Explanation', *Philosophy of Science*, **51**, pp. 212–41.

SOBER, E. [1984b]: *The Nature of Selection: Evolutionary Theory in Philosophical Focus*. MIT Press.

VAN FRAASSEN, B. [1980]: *The Scientific Image*. Oxford.

VON WRIGHT, G. [1971]: *Explanation and Understanding*. Cornell University Press.

Reviews

LUCAS, J. R. [1984]: *Space, Time and Causality*. Oxford at the Clarendon Press. Pp. viii+206. £19.50. (ISBN 0-19-875057-9).

In his preface Mr Lucas says that the book is based on his lectures to first year Oxford students reading Physics and Philosophy, or Mathematics and Philosophy. No special knowledge of relativity or of quantum mechanics is required. Indeed the book itself provides a lucid picture of certain important mathematical ideas such as order relations and equivalence classes, group theory, the real number continuum and connexity.

The volume grew on me as I read it. The author deals very interestingly with those aspects of space and time which can be argued for on the basis of various simple postulates, such as continuity, translation and rotation indifference, and the unlikelihood of action at a distance. Although much of this is elementary, there is a wealth of comment and illustration which is far from being familiar.

The author sees the postulates in question as being closely bound up with causality. No doubt it was for this reason that he deals with causality in the early chapters, before coming to space and time. Useful discussions of induction and modal logic are followed, in Chapter 4, by an ample account of the Millsian treatment of 'cause' and 'effect'.

What I felt was lacking was some discussion of the limits of application of the non-probabilistic version of causality which is used almost throughout the book. (In fact this issue is not mentioned until the last two pages where the author writes briefly about the need for probabilistic causality in the fields of thermodynamics and quantum theory.) Lucas makes it clear (pp. 65–6) that he is not a determinist, but in that case I think he should distinguish, as in Schlick [1961], between 'causality', the existence of laws, and the so-called Principle of Causality, the assertion that *everything* in the world occurs according to laws. Born [1949] makes a similar distinction between causality and determinism. Determinism, as I understand it, asserts a *precise* connection between total states-of-affairs, whereas causality is compatible with a merely statistical relationship between individual 'events'.

Is it the case perhaps that the faith in a Yes/No form of causality had its historic origin in the fact that man had to make his tools from materials which are *rigid*? When he operates on the world with things such as spades and levers, he can expect a univocal result. But imagine a world consisting of liquids and gases, not solids. (This is not merely a possible world; it is the *actual* world of simple organisms living in an aqueous medium!) If man lived in such an environment, determinism might have seemed much less plausible since nothing could be relied upon.

If I have dealt with these points at some length it is because causality plays such a big part in the book. But its author draws back (p. 166)—and wisely, I think—from proposing that time should be *reduced* to causality. As Mehlberg ([1980], vol. I, pp. 113*ff.*) put it, the *local* time concept is based on 'before' and 'after' as the ordering relation; the notion of causal connectibility need be brought in only for the purpose of extending the time concept for application to astronomical distances.

Chapter 5 is largely taken up with the question whether space and time should be treated as discrete, dense or continuous, and arguments are put forward in favour of their continuity. Chapter 6 provides an excellent analysis of the process of measurement, based on ordering relations and the notion of equivalence classes. Useful logical points are made with which many scientists may be unfamiliar. Elementary group theory is introduced in Chapters 7 and 8, and it is shown that the assumed absence of any causal consequences of merely translating a body, or of reorientating it, indicates that space is homogeneous and isotropic, and that time is homogeneous.

The later chapters, 9 to 12, are mostly taken up with the relative merits of relational and absolute views of space and time. The author is inclined to the view that the balance of the arguments (particularly Nerlich's well-known argument about 'handedness') favours the absolute concept of space. Indeed he adduces reasoning (pp. 146, 149, 178) to the effect that space is a philosopher's 'substance'.

Interestingly the same is not said about time! Indeed I don't see how it could be, since one of the prime attributes of 'substance' is that it continues over time (p. 120). Thus, if it were said that time is a 'substance', this would amount to the vacuous statement that time continues over time. But is it consistent with the 3 + 1 notion of space-time, for space to be a substance but not time?

Very little is said about temporally asymmetric processes in nature, but Lucas does accept (p. 150) that time has a quality of 'directedness' and this again he links to causality. It is surprising too that there is no mention of the ontological status of the three-fold division into past, present and future. Other recent books by philosophers of time have also avoided this issue—as if it were in some sense disreputable! Is it due to a tacit acceptance of the idea that the division is purely psychological? This is a tenable position (Denbigh [1981]), but one would wish to see it discussed.

Reverting to my previous remarks about causality, I think there should also have been some discussion of the consequences, for the author's spatio-temporal arguments, of adopting a probabilistic form of causality. Not to accept that events have probabilities over the whole range from 0 to 1 is contrary to the spirit of the principle of continuity which is applied to space and time. Can one have it both ways? The whole business of where, in nature, one may expect discreteness and where one may expect continuity is still profoundly puzzling. Or to put the problem rather differently: Is the concept of infinity really essential to physical science, as distinct from being a mathematical convenience?

But enough of these cogitations! This is a book with which one may disagree in parts (as is so often the case in regard to space and time) and yet find it amply rewarding.

KENNETH DENBIGH
Chelsea College

REFERENCES

BORN, M. [1949]: *Natural Philosophy of Cause and Chance*. Oxford: Clarendon Press.
DENBIGH, K. G. [1981]: *Three Concepts of Time*. Springer-Verlag.
MEHLBERG, H. [1980]: *Time, Causality and The Quantum Theory*. Reidel Publishing Company.
SCHLICK, M. [1961]: *British Journal for the Philosophy of Science*, **12**, pp. 177, 281.

BOWLER, PETER J. [1984]: *Evolution: The History of an Idea*. University of California Press. Pp. xiv+412. (Hardback $29.95) (ISBN 0-520-04880-6).

This book is intended as an introductory survey of the history of evolutionary theories and their social and other implications, particularly as a text for university courses in the history of science, but also as a concise guide to historians. It includes a 55-page bibliography of accessible primary sources and an up-to-date listing of a wide range of secondary sources on this topic. The book succeeds admirably in achieving its aims, and it would also be most useful as a source book and guide to this historical literature for readers of this journal.

Bowler offers a predominantly conceptual history of evolution, and he makes insightful and important points regarding historical variations on key evolutionary concepts, such as those of taxonomical relation and succession, species, variety and inheritance. Little attention is paid to biographies of the persons discussed or to the growth of scientific institutions relevant to the conduct or support of scientific research, though 'external' relations such as social origins of certain concepts and the ethical and social ramifications of biological theories are discussed with attention to a balanced presentation of various viewpoints. Conceptual history illuminates the course of the history of science in a way especially interesting to the philosopher because, in addition to being influenced by the practical and empirical success of these enterprises, conceptual systems support certain scientific enterprises and exclude others. This heuristic role of conceptualisation in scientific development is well illustrated in the various developments Bowler has chosen to describe.

What is unique about this work is its scope and balance. Like others, such as Greene [1959], Bowler describes developments of views that are epistemically auxiliary to evolutionary biology in the fields of cosmology and geology, as well as the history of certain concepts central to taxonomy and natural history, such as linear, multilinear and branching patterns, with or without hierarchical, anthropomorphic or other value-laden interpretations, and in historical or temporalised *versus* idealistic versions (see especially Chapter 3). Especially helpful in regard to taxonomy is the inclusion of well drawn and captioned diagrams. Like Ruse [1974], but with more detail regarding social and philosophical sources, Bowler discusses the social, philosophical and religious context in which Darwin found himself in England, as well as the continental idealist tradition in taxonomy, comparative anatomy, paleontology and goal-directed evolution, and its contribution to, in spite of its incompatibility with, Darwinism (Chapter 4). Also included is an extended treatment of the neo-Lamarckian movement appearing late in the nineteenth and early in the twentieth centuries, which, according to Bowler, was motivated by moral objections to the cruelty of natural selection and a lingering preference for final causes in nature, was not hindered by Weismann's germ plasm theory, and was never unequivocally supported by experiment (Chapters 8, 9). Bowler's book is further distinguished by moderate but plausible externalist historiographical position (Chapter 6), an insightful chapter on the social and ethical implications of Darwinism (Chapter 10), an account of the evolutionary synthesis of the 1930s and 1940s which solved many of the problems either known to Darwin or raised by Lamarckian and saltationist rivals to Darwinism (Chapter 11), and finally recent developments such as further Lamarckian, punctuationist, cladist and creationist challenges (Chapter 12).

Bowler thus covers a lot of ground. Clearly much that is useful and relevant must be left out, particularly the citation of primary sources and other details such as the now well known succession of search, trial, self-appraisal, reformulation and synthesis in Darwin's 'Notebooks on the Transmutation of Species'. However, some of these omissions are at the expense of vagueness and others obscure methodologically significant points about the history of evolutionary theories.

Some of Bowler's claims, particularly concerning the social origins of Darwinism, on first impression seem to be both vague and unsubstantiated. For example, Bowler claims that social thought in early-nineteenth-century Britain was patently influential in the development of Darwin's ideas on natural selection even though the problems he addressed himself to were internal to natural history and geology (pp. 143*ff*.). Bowler tells us, rather vaguely, that 'the individualism of the *laissez faire* school almost certainly is reflected in Darwin's decision to treat biological species as a population of diverse organisms rather than copies of an ideal type' (p. 95), which he claims is evident from Darwin's wide reading in the political and economic literature of his day (p. 145). The relation of 'influence' between Darwin's social philosophy is much clarified in the secondary literature cited by Bowler, particularly in Schweber's claim [1977] that Darwin's interest in instincts, habits and emotions in animals, on the one hand, and in the evolution of a stable moral order for humans, on the other, generates support for a metaphysical view that places man in the evolutionary process (*cf.* Darwin's 'Notebooks on Transmutation', B131, C76–80, C154). Thus, in place of the vague 'influence' one can identify a specific relation of relevance between Darwin's general evolutionary interests and the study of the social environment and its relation to the individual. In this light the connection between social thought and Darwin's evolutionary thinking is one of analogical inference: If there is no essential difference between humans and other animals, particularly the apes or others manifesting seemingly mental attributes, then characteristics of individual activity bearing upon the development and maintenance of a social and moral order in the former ought to be found bearing upon the generation and maintenance of adapted breeding populations in the latter as well. Given the metaphysical assumption of continuity between humans and non-humans, the analogy might rationally warrant Darwin's seeking something corresponding to Adam Smith's individual self interest in non-human organisms, and the same line of reasoning would warrant the importation of Malthus' population pressure and struggle against disease, famine, *etc.* into biology. However, Darwin only mentions Adam Smith in the 'M Notebook', and there is no explicit summary or discussion of Adam Smith's individualism in either the 'M Notebook' or the 'Notebooks on Transmutation'. Hence Bowler's claim about Adam Smith's influence lacks the explicit substantiation available for the common claims about the influence of Malthus on Darwin's search for a theory of species transmutation.

Bowler suggests, with others such as DeBeer and Ghiselin, that Darwin's discoveries are products of the 'hypothetico-deductive method' (pp. 14, 146, 202). However, he seems unaware of the tension between this concession and the account he gives of Darwin's research in the years 1836–9. Although he acknowledges the contributions of Darwin scholarship during this period, he offers very little specific detail regarding the content or the development of the ideas recorded in the 'Notebooks on Transmutation' or in the records of Darwin's field work in South America. Hence there is little specific documentation regarding the 'logic' of Darwin's theoretical discoveries. Darwin's early commitment to the view that speciation must be through the process of natural reproduction is incompatible with hypothetico-deductivism, according to which scientific innovation is achieved through the substantially (though not logically) unconstrained production of trial hypotheses to be tested only against empirically established facts. Darwin's rationale for choosing problems and strategies in his search for a theory of speciation (his logic of discovery) and the arguments he offers in its defence (his logic of justification) hardly fit Popper's and Hempel's model for the discovery process. Rather, evidence especially from the 'Notebooks on Transmutation' and the 'M' and 'N' notebooks on 'Man, Mind and Materialism' supports the view that his search was for explanatory patterns of a prescribed kind, particularly patterns presupposing continuity between human and non-human species and appealing to variations produced through reproduction and preserved through various isolating mechanisms, such as geographical and behavioural isolation. Darwin's arguments against the intelligibility of natural theological accounts of speciation are significant as a first step in the preparation of the abductive arguments, *i.e.* arguments to the best explanation, that Darwin will use both for working hypotheses and for the final product of his search, the theory of evolution by natural selection. Bowler implicitly acknowledges the import of the argument from explanatory power (abduction) in the reception of Darwin's theory (p. 178), though this important form of scientific argument is never mentioned in Bowler's methodological discussions. Interestingly, in his excellent appraisal of recent 'scientific creationism', Bowler offers just the arguments that both Darwin and Wallace offered against natural theological theories of speciation and in favour of their position regarding the kind of explanation to be sought for speciation, *viz.* arguments appealing to the superior epistemic accessibility, coherence with theories in other relevant domains, and explanatory adequacy and the relative lack of arbitrariness in the naturalistic explanatory scheme (pp. 344*ff.*).

Bowler also clearly delineates the methodological deficiencies of natural selection in the light of mid-nineteenth century knowledge and observational capability. The theory conflicted with Kelvin's highly credible (but false) thermodynamic studies of the rate of the earth's cooling, thus raising a problem in which the evidence from external incoherence is in conflict with geological evidence from stratigraphy and erosion (p. 194). Nageli had pointed out that many of the characters distinguishing closely related species had no apparent adaptive value (p. 196). Mivart raised questions about the adaptive value of intermediates in the selective process by which functional organs were supposed to have developed (p. 197). Darwin's eventual commitment to sympatric variation made him vulnerable to Jenkin's argument that such variants should blend out by crossing with non-variants in the population (p. 199). Darwin was unable to provide a cause or a suitable mechanism for the production or transmission of variant characteristics (p. 204). Natural selection is not exemplified in any immediately observable process, and hence the theory lacks observational 'proof' (p. 200). Some historians have suggested that Darwin became increasingly aware that natural selection cannot be reconciled with traditional theology, even the deism that would allow God's benevolence to be achieved indirectly through natural laws (p. 208). The absence of any defensible notion of progress in the successive products of evolution undermined the view that the sacrifice of individuals in the struggle for existence is for an overall good. This produced tensions not just between evolutionists and theologians of the latter part of the nineteenth century, but also in the minds of scientists who found Darwin's materialism 'distasteful' or morally objectionable, and disposed these scientists to reconsider teleological evolutionary mechanisms such as Lamarckism (pp. 208, 214, 244).

These seemingly legitimate objections should focus the interest of the philosopher on what, if any, methodological virtues natural selection had which would warrant the tenacity of the Darwinians, tenacity which in retrospect was successful if not rationally justified (*cf.* Feyerabend [1974]).

The eventual success of the 'evolutionary synthesis', as Bowler presents it, can be viewed as solving or contributing to the solution of the above mentioned problems facing the original form of natural selection theory. These developments are exactly what one would expect from a research programme deliberately directed at eliminating the epistemic deficiencies of Darwin's theory, although Bowler does not provide the kind of evidence that would establish the particular objectives of the authors of the synthesis. The multiple gene hypothesis, the mathematical studies of the spread of advantaged genes through and the maintenance of genetic variation within breeding populations, new support for isolation of subpopulations as a source for evolutionarily significant variation, and statistical methods in paleontology answered some of Darwin's questions about the causes of variation and brought to bear the highly productive field of Mendelian and population genetics, as well as molecular biology and paleontology, as evidence that Jenkin's objections could be answered (pp. 289–300).

In sum, the few shortcomings of this book in clarity and evidence and in the discussion of methodology are far outweighed by its scope and overall adequacy to the subject undertaken. The final section on creationism should be required reading in all introductory philosophy classes, especially in the 'bible belt' of the USA!

SCOTT A. KLEINER
University of Georgia

REFERENCES

FEYERABEND, P. K. [1974]: *Against Method*. New Left Books.
GREENE, JOHN C. [1959]: *The Death of Adam*. Ames: Iowa State.
RUSE, M. [1979]: *The Darwinian Revolution*. University of Chicago.
SCHWEBER, S. [1977]: 'The Origin of the *Origin* Revisited'. *J. History Biology* **10**, pp. 229–36.

MAXWELL NICHOLAS [1984]: *From Knowledge to Wisdom: A Revolution in the Aims and Methods of Science*. Basil Blackwell. Pp. vi+298 (Hardback £19.50.) (ISBN 0-631-13602-9.)

We all know that the old triumphalist faith in Science is now in retreat. That previous certainty, that the methods and results of natural science could be adapted and applied to solve all mankind's problems, is lost. We are not even sure that Science can solve the problems that the science/technology system itself creates. The veridical character of scientific 'facts' becomes ever more difficult to justify or even to define. And these changes are not merely new difficulties of puzzle-solving within academic philosophical paradigms. In very many ways, our secular civilisation depends on a symbolic image of Science, the descendant of the Enlightenment, anticlerical vision, at its core. Should that weaken, the consequences for the ordering of our education, and of all our social institutions, might be slow in manifesting but they would certainly be deep.

One might say that the crisis at the philosophical level has been patent ever since the young Popper boldly sacrificed the True of Science in order to preserve the Good, through 'Falsifiability'. Decades have gone by; Kuhn was supremely indifferent to the issue and so instantly made the crisis much worse; Lakatos attempted a rescue-job, and Feyerabend played the Latter-day Dada. What then? Can anyone be a new Neurath of the 1920s' *Wienerkreis*, now that so much of oppression and evil wears a scientific, not a clerical face? If not, what way forward can there be?

To be sure, there are many prophets of the Journey to the East that started in the '60s. But, however much one may find uplift in the Tao of this or the ecology of that, the message there is for so complete a transformation of our reality, as to be beyond mere good will and education. Is there, then, some way to a reform of Science, within that style of Western rationalism in which we all do our thinking?

This, I believe, is the problem-situation for Nick Maxwell. He is very much a philosophers' philosopher, depending on reasoned argument as the pivot of his proposed revolution. Indeed, one can see him as a sort of 'left-Popperian', drawing on Popper's critical insights and extending them to the whole closed society of contemporary incorporated science.

As befits a philosopher, he defines his key terms clearly, and then uses them consistently. What he calls 'the philosophy of knowledge' refers to the dominant accepted aim of intellectual inquiry: that it should provide objective, factual knowledge, considered as dissociated from any human concerns that may have motivated or conditioned the inquiry. This 'philosophy' is most clearly expressed in what Maxwell calls 'standard empiricism', the guiding philosophy of all inquirers that would be 'scientific'.

In striking contrast to this is the 'philosophy of wisdom'. By this Maxwell does *not* mean the pursuit of some unspecifiable knowledge beyond mere understanding. He is quite precise about his 'philosophy of wisdom'. It is best to quote him: 'The central and basic task of rational inquiry . . . is to help us imbue our personal and social lives with vividly imagined and criticized possible actions so that we may discourse, and perform, where possible, those actions which enable us to realize what is of value—happiness, health, sanity, beauty, friendship. . . .' Wisdom, then, is about value, including of course the value in knowledge itself. Its appropriate methods are diametrically opposed to those of the 'philosophy of knowledge' people as beings who feel, imagine, explore, and co-operate, are the agents of this inquiry; and its results are primarily and explicitly intended to enhance awareness, love and life.

The author has great respect for the ideals and the achievements of 'the philosophy of knowledge'. But he sees it as fatally flawed, in the rejection of 'wisdom' from its purview. He remarks in passing that such a tame, instrumentalist science can flourish even under a totalitarian regime; something that should produce a pause for reflection to those who have traditionally associated 'science' of just this neutralised sort with 'democracy'. Further, he argues that 'the philosophy of knowledge' is self-destructive in that the exclusion of speculations, values and purposes renders the process *irrational*; Kuhn's description of science alternating between myopic puzzle-solving and totally blind paradigm-shifting holds only too well. The rejection of 'wisdom' even among philosophers of science means that philosophy itself is reduced to an automatically self-correcting, anti-critical state.

Operating within his 'philosophy of wisdom', he undertakes his philosophical task with a very practical end in view. The world of people is in an appalling state; and to remedy this we need not primarily more knowledge but rather new ways of *acting*. He does *not* mean that knowledge should be neglected or despised, in fact, he believes that it will be more fully appreciated, even the most 'pure' of the sciences, and only when the human dimension of its production is understood. But we must have a new tradition of inquiry to help us solve, not ignore, our problems of *living*; otherwise this 'major intellectual disaster at the heart of western thought' will continue to incapacitate us.

All this may sound rather general in its scope; and indeed most of Maxwell's argument is at that level. But there are some illustrative quotations that make his argument very difficult to dismiss. The pre-war scientistic optimism of a Baroness Wootton, who simply could not see why mankind should be different from atoms, as an object of knowledge, may be by now a familiar relic. Less familiar is the recent paper on 'Physics and Modern Warfare—The Awkward Silence' by E. Wollett, remarking on the absence of public discussion among physicists of the many intimate links between physical science and weapons-systems development. Nick Maxwell himself supplies statistics: of 109,577 contributions recorded in *Physics Abstracts* in 1980, those devoted to 'philosophy of science' and 'science and society' number 38 and 14 respectively. And none of these are really critical anyway!

Suppose, then, that a major source of priorities, funds and graduates' jobs in 'pure science' is military; and that this aspect of science is deemed irrelevant by the overwhelming majority of those who research, teach, sociologise, philosophise or moralise about science. What are we to make of such a phenomenon? It is in part a *political* situation, in its causes and effects; but it is also *philosophical*, and this is Nick Maxwell's point of focus. Such a gigantic co-operative endeavour of concealment, amounting to a huge deception, could be accomplished naturally by all the educated, humane participants, a 'conspiracy needing no conspirators', *only* because their 'philosophy of knowledge' envelops them in the assurance that their directors, paymasters and employers have nothing to do with the *real* thing—the research.

This, to me, is the heart of Maxwell's message. It recalls Popper's remark that Kuhnian 'normal science' is a danger to science, and to civilisation. There is much to criticise in this book; indeed, something for everybody. Those who see the problem mainly in historical/political terms (like myself) can become impatient with his intellectualistic analyses of causes and remedies. Those who see value only in refined conceptual analyses will be irritated by his insistent, didactic style of argument. But that is missing the point of the book, and its real contribution. How much longer can we honestly philosophise about natural science while implicitly denying the existence of the Bomb, A-, H- and Star Wars too? Nick Maxwell has raised the challenge, constructively and from within the philosophy of science. He deserves our appreciation.

JEROME R. RAVETZ
University of Leeds

GOOD, I. J. (1983): *Good Thinking: the Foundations of Probability and its Applications*. University of Minnesota Press. Pp. xviii + 332.

This book is a selection from Good's very considerable output of papers in philosophy of science (considerable both in importance and in cardinality), and in particular in "quantitative" philosophy of science, or what he calls the mathematics of philosophy. Many of the papers are gutted of their mathematical details, and to some are added later comments.

These papers are classified in the book under five broad headings: (i) Bayesian Rationality; (ii) Probability; (iii) Corroboration, hypothesis Testing, Induction and Simplicity; (iv) Information and Surprise; (v) Causality and Explanation (I use Good's own section headings). Many topics, however, are discussed under more than one heading, which is inevitable in view of the strong interdependence, within Good's particular philosophical scheme, of probability, corroboration, information, rationality and explanation. It is not possible within the span of a relatively short review to give this book the sort of extended discussion which, as a summary of the philosophical contributions of one of the most distinguished living contributors to the philosophy of science, it deserves. I shall instead give a brief outline of the principal topics of each of the sections of the book, with, where I think it called for, some comment.

In the papers of the first section the reader is presented with Good's conception of rationality. This involves a subjective probability distribution over a class of possibilities, including a class of theoretical hypotheses, though the probabilities that we actually use are in general, according to Good, only interval-valued, which means in effect that only what are called upper and lower probabilities are assigned to the events concerned. Point-valued probabilities nevertheless play for him an important role in being useful idealisations, and Good presents in several of the papers in this book an inferential model in which probability-inequalities are fed into a "black-box" incorporating the classical (point-) probability calculus, and in particular Bayes's Theorem, yielding derivative inequalities as outputs. The initial inequalities can be derived from inequalities between upper and lower probabilities, which Good introduced in a novel way in his 1960 paper "Subjective Probability as the Measure of a Nonmeasurable Set" (reprinted here in section (ii)), by defining the upper (lower) probability $P^*(E/F)$ ($P_*(E/F)$) to be the greatest lower bound (least upper bound) of numbers x for which "You" can discern that $x > P(E/F)$ ($x < P(E/F)$), on the usually counterfactual assumption that $P(E/F)$ exists. From an inequality of the form $P^*(E/F) < P_*(H/F)$ you can then infer the inequality $P(E/F) < P(H/F)$; clearly, the inequalities so obtained amount only to a partial ordering of the P-terms except in the degenerate case where all the upper and lower probabilities coincide (they do of course inside the black box).

Similar considerations apply, for Good, to utilities: actually, we do not have sharp utilities, but it is often a convenient and useful idealisation to suppose that we do. Good's theory of rational behaviour asserts that one should behave so as to maximise expected utility per time-unit (the qualification is important since if the cost of reasoning, for example, is not negligible then your assessment of the probabilities and/or utilities involved may differ from what they would be on the fullest consideration). Utility is taken by Good to be anything that is of value in the context and whose value will in general vary depending on which possible contemplated situation will be realised. For this reason he often uses the term "quasi-utility": examples of quasi-utilities whose expectation one might in appropriate circumstances wish to maximise are information, in either Fisher's or Shannon's sense, explanatory power (of which he gives a formal explicatum in the last paper in the book), simple monetary profit, weight-of evidence (when considering which type of experiment will maximally discriminate between competing hypotheses), etc.

The precept of what Good calls type II rationality, namely to call a halt to further deliberation when the costs involved have risen sufficiently high, means also in effect *not* always obeying the classical probability calculus, for the latter embodies axioms with antecedents like "if $E \vdash H$ then" or "if $E \vdash \bar{H}$ then". To be able to know with any definiteness that the relation $E \vdash H$ holds or not for arbitrary E, H is a problem for which there is known to be no algorithm, and in practice is an unrealistic demand. Correspondingly, Good suggested many years ago, and again here, most notably in the paper "Dynamic Probability, Computer Chess, and the Measurement of Knowledge", amending the axioms of subjective probability in such a way that the relevant ones are now prefixed with "if you know that $\mathrm{E} \vdash \bar{\mathrm{H}}$ then". The notion of probability so generated he calls "dynamic or evolving probability" since you may well change your probabilities on receipt of logico-mathematical information only. Clearly, some applications of probabilistic-inductive reasoning to logic or mathematics—assessing, for example, support for a generalisation about all natural numbers based on sampling the first n—would require a weakening, of this type, of the usual calculus (if Peano's axioms were taken as background data and the generalisation was as a matter of unknown fact a theorem). My own feeling is, however, that this weakening is undesirable, for the reason, which Good endorses, that the subjective probability calculus is held up as an ideal of reasoning, like deductive logic, and therefore one would have thought that it is explicitly not required to mirror human fallibility. If the calculus is amended in the way Good proposes then such arguments for its prescriptive status as the Dutch Book Argument no longer work, since somebody with better deductive powers will be able to make risk-free money out of "dynamic probabilists".

In his discussion of probability, in the second section of this book, Good affirms in his 1959 paper "Kinds of Probability" the existence of physical probabilities, though here and elsewhere he claims that they can be measured only with the help of a theory of subjective probability. He points out that frequency definitions of probability like von Mises's run into problems of empirical checkability; however, he does not himself offer any systematic account of physical probabilities. He does, nevertheless, asse[illegible] (pp. 71, 72) that physical probabilities in a deterministic environment must necessarily be given a subjective interpretation (so *qua* physical probabilities they do not really, or Really, exist), since they must be attributed to our own ignorance of the fully determining causes of the events. This is boldly stated, and so stated seems to be a *non sequitur*: the statement that a certain type of set-up has a measurable tendency to produce events of type A does not by itself appear to imply anything about the existence or non-existence of a set of fully determining causes of those events.

Good distinguishes several subtypes of subjective probability in this same paper, among which there is what he calls logical probability, regarded by him as an ideally objective subjective probability, or "credibility", if there is one, and not, as with the early Carnap, as something laid down totally *a priori* for types of formalised language (see also his more extended discussion on p. 151).

In his third section, on the methodology of testing and confirmation, Good advances what he calls his compromise between Bayesianism and non-Bayesianism. His avowal of dynamic probabilities, and an insistence that subjective probabilities are not in reality totally ordered, and so are not numbers, already distances him from full-blooded Bayesianism, as does his endorsement of some standard non-Bayesian inferential techniques, like the use of tail-area probabilities, and of the less standard weight-of-evidence function (the weight of evidence in favour of H provided by E given background data G he defines to be, where "o" means odds,

$$\log(\mathrm{o}(\mathrm{H}/\mathrm{E} \wedge \mathrm{G})/\mathrm{o}(\mathrm{H}/\mathrm{G}))$$
$$= \log(\mathrm{P}(\mathrm{E}/\mathrm{H} \wedge \mathrm{G})/\mathrm{P}(\mathrm{E}/\bar{\mathrm{H}} \wedge \mathrm{G}))$$

given the usual relationship $\mathrm{o} = \mathrm{P}/(1-\mathrm{P})$. But he is careful to insist that the non-Bayesian techniques are to be used only where prior probabilities can in effect be neglected: in other words, their techniques are to be used only in circumstances in which a full Bayesian analysis would yield the same judgements. But as Good points out, given the typical indefiniteness of people's initial probabilities, even the committed Bayesian is forced to rely on non-Bayesian techniques as the upper probability approaches 1 and the lower approaches 0; or as he puts it himself, all non-Bayesians are Bayesians with upper initial probabilities of 1 and lower initial probabilities of 0 ("To the Bayesian", he remarks, "all things are Bayesian"). Scattered throughout the papers in this book are familiar but none the less penetrating criticisms of non-Bayesian techniques which are unilluminated, as it were, by a Bayesian awareness of their limitations: the problem of optional stopping, for example, is raised, and the standard response, that one should formulate criteria independently of the data, is shown to entail the possibility of drawing equally misleading conclusions from the data.

The fourth section of the book contains Good's classic 1967 discussion "On the Principle of Total Evidence", where he showed that the expected value relative to a further observation, of the maximum of the expected utilities of a set of acts, exceeds the maximum of the expected utilities taken relative to the initial probability distribution. Thus he demonstrates a practical justification, in terms of a rationality principle based on maximising expected utilities, of the principle of total evidence. (Another justification is provided by Shannon's result that the expected decrease-of-uncertainty information about the value of one variable obtained by observing the value of another non independent variable is always positive.) Also in this section is the paper "The Probabilistic Explication of Information, Evidence, Surprise, Causality, Explanation and Utility" (1971), in which there is a discussion of the relation between various criteria for selecting prior distributions and the minimisation of a type of expected utility. There is also a discussion of Miller's paradox, though Good's solution to it, like many others, does not I think succeed in locating in it an uncontroversial logical fallacy (Graham Oddie and I have argued in this journal (vol. 30 pp. 253–278) that it does involve a quite "standard" logical fallacy, namely, substituting a term which is not free for the variable substituted).

The final section contains papers devoted to the explication of "the tendency of an event F to cause an event E" and of the notion of the explicativity of a hypothesis H with respect to evidence E. In "A Causal Calculus" Good shows that up to a continuous increasing transformation the constraints he lists on an adequate explicatum determine Q(E : F/G), "the tendency of F to cause E conditional on G", to be measured by the weight of evidence against F if E does not occur *i.e.* by $\log(o(\bar{E}/\bar{E} \wedge G)/o(F/G)$ (note that if E and F are independent given G then this quantity goes to 0). In the final paper "Explicativity: A Mathematical Theory of Explanation with Statistical Applications" the details of the applications are unfortunately omitted. Good proposes a quantitative explicatum of the explicativity of H with respect to E as $\log(P(E/H)(P(H))\gamma/P(E))$ where $0 \leqslant \gamma \leqslant 1$, which with $\gamma = 1$ gives $\log P(H/E)$, and with $\gamma = 0$ gives what earlier good and Hamblin had independently offered as an explication of "explanatory power". Good points out that with $\gamma = 0$, adding irrelevant and possibly absurd clauses to a hypothesis which already predicts E makes no difference according to this measure: it stays at $-\log P(E)$. If $\gamma = 1$ is chosen then a tautology has maximum explanatory power. For these reasons Good places $\gamma = 1$ is chosen then a tautology has maximum explanatory power. For these reasons Good places γ strictly between 0 and 1, and points out that it can be regarded as an index of how desirable we think the use of Occam's razor is.

This is a valuable collection; it contains a wealth of informed and acute *obiter dicta*, on Fisher, Popper, Jeffreys, Carnap, besides presenting an exposition of Good's own philosophical views. This reviewer, for one, feels after reading these papers that the case for making explicit reference to prior probabilities—that is to say, for accepting the basic Bayesian premise—rather than sweeping them under the carpet, has never been made so persuasively ("sweeping under the carpet" is seen by Good as a methodological position of such popularity that it is afforded acronymic status by him, as SUTC). As is to be expected, there are also some good jokes in this book. Finally, the author (I. J. Good) wishes to point out that the word "not" has been omitted from the expression "when you do know it" on p. 54, line 2 up.

COLIN HOWSON
January, 1986, L.S.E.

NIINILUOTO, ILKKA 1984: *Is Science Progressive?* Reidel. *xi*+272 pp.

In this collection of essays Niiniluoto defends a realist view of science, and explores its ramifications for the natural and social sciences, philosophy, mathematics, operations research, and technology. Seven of the twelve essays are reprinted, but only three of the twelve have been readily available in English.

The scientific realist holds that truth is the primary goal of scientific inquiry. If there is to be scientific progress, it must be progress towards the truth. And if progress can be gradual, then closeness to the truth must admit of degrees. The old true/false dichotomy need not be rejected, but must be supplemented with a coherent account of closeness to the truth, which satisfies the requirement that scientific progress at least be possible (p. 6). Without such an account realists must either abandon their account of the aim of science (and hence give up their realism) or abandon the idea that progress admits of degrees. Unfortunately the problem of truthlikeness is rarely acknowledged by modern advocates of realism. Niiniluoto's collection of essays gives the problem its due weight, and is therefore a welcome addition to the literature.

It is the work of Popper which has brought truthlikeness such attention as it has received, but Niiniluoto shows that this is just one more of the many features of Popper's philosophy already foreshadowed in Peirce's. In tracing the development of the realist thesis from the nineteenth century (chapters 3 and 4) Niiniluoto argues that Peirce occupies a central and largely unacknowledged position in the realist tradition. From the Peircian thesis that science is a 'self-corrective process' Niiniluoto draws what he takes to be the fundamental features of science: objectivity, critical attitude, autonomy, and progress (chapter 1). Some of his pronouncements (like 'reality depends upon the ultimate decision of the community' (p. 49)) threaten to push Peirce well out of the realist camp, but Niiniluoto does not shrink from explaining such pronouncements within his realist interpretation, and the result is both illuminating and highly plausible.

In the central essays of the volume (chapter 5, "Scientific Progress", and chapter 7, "Truthlikeness, Realism, and Progressive Theory Change") the broad outlines of an adequate theory of truthlikeness, based on Hintikka's distributive normal forms, is sketched. A measure of distance between maximally strong propositions (or their surrogates, constituents) can be made to yield a theory of distance from the truth. To my mind this is by far the most fruitful approach to the problem.[1] I am less enthusiastic about Niiniluoto's proposed liberalisation of this program. In chapter 7 it is argued that Hintikka's constituents are just one of many different ways of cutting up the logical space, and different distance measures based on these different ways are suitable in different contexts (pp. 166–7). This liberalisation certainly makes it easier to demonstrate desirable orderings for certain kinds of theory (notably those which involve continuous functions) as is elegantly shown (pp. 172–3). But the liberalisation threatens the original simplicity, unity, and boldness of the programme based on normal forms. All theories can be expressed by means of the normal forms of some sufficiently rich (probably higher-order[2]) language. Consequently

---

[1] I argue this at length in Oddie [1986].

[2] A theory of truthlikeness for higher-order frameworks is developed in Oddie [1986], pp. 122–135, and 191–211.

Niiniluoto's liberalised programme will involve either redundancies or inconsistencies. For any two alternative ways of defining truthlikeness will either converge in their results (in which case one will be redundant) or they will diverge (in which case the espousal of both will involve an inconsistency). There is a danger here of a repetition of an unfortunate feature of the recent history of inductive logic and modal logic: acceptance of a proliferation of theories with no generally acknowledged criteria for choosing the correct one, or even a perceived need to do so. This is the kind of development over which critics of realism might well rejoice.

One common objection to the concept of truthlikeness, that it is of no methodological use, is dealt with at length. The objection goes something like this. There is no infallible way of telling whether a new theory is more truthlike than an old theory. To be able to decide this we would need to know the truth itself, and if we knew that, degrees of truthlikeness would cease to be of interest. Niiniluoto points out that degree of truthlikeness could be adequately *estimated* (like any unknown quantity) provided we had a good system of inductive logic. The best estimate would be the expected value of the degree of truthlikeness, given the evidence. The move from an old to a new theory *is* progressive if the new theory is closer to the truth than the old. The move *seems* progressive if the expected truthlikeness of the new theory is greater than that of the old. Appearances can be deceptive, and future evidence may overturn the judgement of seeming progress. But the realist does not have to embrace infallibilism.

The real difficulty with this solution to the problem of estimating degrees of truthlikeness is the current state of the theory of inductive logic. Multiple continua of 'inductive methods' have blossomed along with the kind of relativism criticised above. With a choice of different measures of truthlikeness, and of different measures of probability, how is one supposed to calculate *the* expected value of degree of truthlikeness?

A different measure (which echoes features of Popper's notoriously inadequate account of truthlikeness) is applied to the problem of progress in mathematical inquiry. Where such an inquiry is focussed on a specific structure (like that of the natural numbers) one theory can be closer to the truth than another simply by implying more truths about that structure. However, this realist account of progress fits some but not all aspects of the growth of mathematical knowledge (pp. 196–200). In other areas, like group theory, something closer to the structuralist view of theories applies. This is not too surprising, because the structuralist view of theories (that a scientific theory is neither true nor false *simpliciter*, but applies to certain structures and not to others) is really just a generalisation to the factual domain of an already established and plausible view of mathematical theories which grew out of the successes of model theory. While the structuralist view does seem to fit some features of mathematical progress, as a view of scientific theories it is explored, and rejected, in chapter 6. There are other aspects of mathematical progress which do not easily fit the realist conception: constructing more elegant proofs, showing that a number of apparently unrelated results can be reduced to a single result, producing more abstract and generalised theories which subsume a number of different theories (pp. 196–200).

The burden of the argument in the final four chapters rests on the distinction between '*cognitive problems* (what to believe, what can be known?) and *problems of action* (what to do?)' (p. 253). Many of the criticisms of the realist conception of the social sciences rest on a failure to make the distinction (pp. 220–1). The finalisation thesis (that at a certain stage of development in a scientific inquiry external social goals set the agenda) challenges the autonomy of science, but the challenge arises from a conflation between the cognitive and the practical (chapter 10). Operations research solves problems of action rather than seeks the truth. Thus the fact that the Kuhnian (and structuralist) conceptions of scientific inquiry fit it well in no way threatens the realist conception of scientific inquiry in general. Finally, technological progress is different from scientific progress, the former being governed by practical utilities and values, the latter by epistemic or cognitive utilities, the overriding epistemic utility being that of truthlikeness (chapter 12).

In the last few pages Niiniluoto takes issue with a kind of attack on the progressiveness of science quite different from those he considers elsewhere: the so-called Romantic challenge to science. The strategy he adopts is the familiar one: we should blame technology for the evils for which the Romantic blames science, the distinction being sustained by the general distinction already mentioned between the cognitive and the practical. Bacon's 'naive technological optimism' (science brings technological progress, which in turn brings social progress) has 'proved mistaken: technological progress alone is not a sufficient condition of social progress' (p. 263). The Russellian thesis, that scientific and technological progress are necessary conditions for general social progress, is endorsed, and it is claimed that this is the rational kernel of Baconian optimism.

This dismissal of the Romantic's case is too swift, and brings to light a weakness in the neat distinction between the cognitive and the practical. For one thing, the distinction suggests that scientists are mere *observers*, gleaning knowledge about nature without intervening into it. But of course, one of the distinctive features of scientific inquiry is the extensive use of experimentation. An experimenter forces an object or system into a state which it would not otherwise have entered in order to find out how it will react. The primary end of the experiment is knowledge, but the experiment involves an action as a means to that end. And so over every experiment there hangs the question of whether or not the cognitive end is justified. It is obvious that certain kinds of knowledge can be obtained only at a cost to more important values. To choose a drastic example, suppose that the only way to obtain reliable information on the nuclear winter theory (that a sizable number of roughly simultaneous nuclear explosions would bring about drastic climatic changes) is to start a nuclear war and see what happens. Non-cognitive values must set limits to our truth-seeking activities.

Once this is conceded the Romantic can develop a case against the scientific enterprise which does not challenge any of Niiniluoto's major theses. He need not challenge the thesis that the primary aim of an inquiry is truth, nor the thesis that scientific and technological progress are essential for the obtaining of some other highly desirable goals. The Romantic need only argue that the *expected value* of engaging in the scientific enterprise is lower than the expected value of not doing so. One powerful argument for this is akin to Pascal's wager. Given actual circumstances (human nature, political realities) the development of nuclear physics makes the development of nuclear weapons highly probable. (Scientists who had no interest at all in the goals of the military might nevertheless develop nuclear explosives for their own experimental purposes, or other purely cognitive ends.) The development of nuclear weapons renders the annihilation of the race more probable than it would otherwise be. (Even if nuclear deterrence works war caused by error cannot be ruled out.) Our annihilation involves non-cognitive and cognitive disvalues on a massive scale (Pascal might say, infinite). Thus even if the probability of nuclear annihilation were small the disvalues associated with it are so great that they are likely to swamp the other factors in our calculation of the overall expected value of the scientific enterprise. If the Romantic critique of science is to be introduced in order to be dismissed then it requires the same kind of detailed careful consideration which Niiniluoto devotes to his other adversaries.

By focussing on the central issues of these essays, the realist conception of inquiry and the importance of the concept of truthlikeness for that conception, the impression may arise that the essays are themselves narrowly focussed. This is far from the truth. Vast stretches of the literature in the philosophy of science are covered, and there can scarcely be a single philosopher of science whose ideas are not discussed somewhere in these pages. Thus the volume is as much a broad survey of the philosophy of science, from a realist's viewpoint, as it is a sustained and vigorous defence of realism.

GRAHAM ODDIE
University of Otago

REFERENCES

ODDIE, G. J. [1986] *Likeness to Truth*. 1986, Reidel.

O'HEAR, ANTHONY (1985): *What Philosophy Is* Penguin Books 316 pp. (Paperback only, £4.95)

O'Hear's book divides into five main sections, on Metaphysics, Epistemology, Logic and Language, Human Beings, and Ethics and Politics. His approach to the topics in each of these areas is more-or-less the same: an initial characterisation of some problems, followed by a survey of currently available theories and solutions. Thus, in the 11 page discussion of truth (which curiously comes under Epistemology and not Logic), there are rapid expositions of the correspondence, pragmatic, coherence, redundancy and Tarskian theories of truth; the 19 pages on naming and referring provide a quick survey of Mill, Frege, Russell, Kripke and Putnam; and so on. Readers of this journal in particular may be interested to know that a 12 page discussion on 'Truth and Relativism in Science' is included in the Epistemology chapter. This raises the question of what justifies our acceptance of scientific theories, takes us through Popperian irrationalism and Kuhnian relativism, and concludes with a brief examination of instrumentalism and realism. One unusual feature for a general introductory text of this kind is a section of some 17 pages introducing propositional and predicate logic. Formalisation is not, however, used elsewhere in the text. The framework of all the discussions is resolutely analytic. Although Sartre gets a passing mention, and Hegel and Nietzsche a couple of references each, the tradition within which O'Hear operates is basically that set by Descartes and Hume, as continued by Frege and Wittgenstein and updated by Quine and Davidson, Putnam and Kripke.

The style throughout is lucid and restrained, and should be well-suited to those with little or no background in philosophy for whom O'Hear is writing. But given that that is his intended audience, he is somewhat cavalier about introducing technical terms. Thus the contrast between direct and indirect perception, and between types and tokens, the concepts of speech acts, logical constructions and others, all make their appearance with no explanation at all. Other positions, *e.g.* instrumentalism in the philosophy of science, are presented in only one or two sentences, which scarcely give an adequate characterisation of the theory (let alone a grasp of what can be said in its favour) before it is demolished.

There are also certain oddities from a philosophical point of view. In his discussion of human beings, for example, O'Hear argues that a functionalist approach, while in many ways attractive, can account neither for sentience nor for free will. But he then offers us no rival account of sentience, nor even enough of a characterisation of it to explain why any functionalist theory, even some improved one of the future, is doomed never to accommodate it. Free will *is* further discussed, but what O'Hear says about it leaves it again unclear why it should present a problem for functionalism. He simply asserts without argument that determinism would leave us all entirely 'passive'—a claim which is all the more surprising, since he has earlier accepted a broadly regularity view of causation.

Again, doubts are raised by his discussion under Metaphysics of the existence of physical objects. He argues that physical objects must be irreducible entities on the ground (itself epistemological rather than methaphysical) that unless we believed that there were such entities, we could 'neither classify our experiences nor break out of them' (p. 34). It is not altogether clear what this claim, with its unhappy metaphor, means. Perhaps the most plausible reading would be that it makes the twin assertions that (1) 'X can describe his experiences' entails 'X believes that there are physical objects', and (2) 'X has knowledge (or perhaps just rational belief?) about something other than his experiences' entails 'X believes that there are physical objects'. Neither of these claims, which in turn can be given several alternative readings, would show that physical objects exist, let alone that they have irreducible existence. In support of them, O'Hear invokes a version of Wittgenstein's anti-private-language argument. But what his discussion then fails to explain is how, *if* there is a problem about how to 'classify' or 'break out of' experiences, it is solved if and only if one believes that there are physical objects. It seems on the face of it that subjunctive-supporting generalisations about experience would be just as effective in solving problems about the solipsist's identification of his experiences, as beliefs in physical objects are alleged to be.

It is interesting to see how relatively little the topics included in introductory texts have changed over the last quarter of a century or so. The area covered by O'Hear is substantially the same as that covered by Hospers in his *Introduction to Philosophical Analysis* of 1956, or Flew in his *Introduction to Western philosophy* of 1971. Of course, differences in emphasis are present (Hospers and Flew took the philosophy of religion more seriously than does O'Hear; O'Hear takes referring and meaning to be more important than did the earlier writers). And O'Hear also provides what is missing in Hospers and Flew and also in more recent introductions, such as Hollis's *Invitation to Philosophy* of 1985, namely a concise statement of the views of contemporary opinion-forming philosophers. Perhaps the text which O'Hear's most closely resembles is, surprisingly, Quinton's *The Nature of Things*, in an updated and more elementary form. For although Quinton was concerned to argue a case, whereas O'Hear is in general concerned simple to present issues and possible solutions, the two works share the same sort of balanced and encyclopaedic approach.

Two small but irksome weaknesses in the book (of a kind which could easily be removed in a second edition) concern the Index and the Further Reading. A random check on the Index revealed (i) that important references to induction (on pp. 62 and 117–124) were not included in the Index, and (ii) that Bernard Williams, whose views on personal identity are usefully discussed on pp. 247–249, does not appear in the Index at all. As regards the Further Reading, given that the main vehicle for modern philosophical discussion is the journal article and not the book, and given that O'Hear's book is intended primarily for students who will presumably have access to the relevant journals, it is a pity that no journal articles (or at least no unanthologised ones) are listed. But these are two minor weaknesses which do not seriously affect the value of the book. It is clear, comprehensive, well-informed and up-to-date, and as such is a welcome addition to currently available introductions to philosophy.

NICHOLAS EVERITT
University of East Anglia

TRIGG, ROGER [1985]: *Understanding Social Science: A Philosophical Introduction to the Social Sciences*. Oxford: Basil Blackwell, viii+224 pp. Hardback £18.50. ISBN 0 631 13365 8. Paperback £5.95. ISBN 0 631 14161 8.

This book seeks to map—with a student audience in mind—the current state of play in the philosophy of social science. It has manifest merits. It is superbly written, being at once lucid, witty and as simple as its subject matter allows. It is exceptionally well informed, and it has a remarkable range—no significant debate in the recent philosophy of social science is omitted, and the positions of most of those who have contributed to such debates is characterised with unfailing fairness and courtesty. Furthermore, the book as a whole has something of a plot, and this is worth spelling out.

Trigg begins by examining current views of the nature of science so as to get purchase on question as to whether there can be a *science* of society. His report from that front is amusingly different from that famously provided by Thomas Kuhn. Philosophers of natural science disagree amongst themselves and are unable to offer any cognitive model, in large part because empiricism has been, in Trigg's eyes, so effectively criticised as to have led to its being abandoned. The claim that science is but one 'practice' amongst others leads naturally to a discussion of the sociology of knowledge. Is there anything to its claim that socioanalysis of particular positions will allow one to arrive at some sort of objective truth? Trigg elegantly dismisses the programme of the sociology of knowledge as self-contradictory. Social science can only make sense, Trigg continues, if it has direct access to its own reality. This leads to a discussion of the debate between those insisting on the importance of agency in social life and those asserting the prior reality of social structure, and this is followed by an exceptionally lucid analysis of problems involved in understanding the beliefs of primitive societies. Trigg insists that there can be no social science if every culture is absolutely its own world: if this were so we would simply have the choice of staying in our society or joining another. That we can translate between societies is possible, in Trigg's view, because every society faces 'the real world' and because its members have a shared biological make up. No particular guidance is offered as to exactly what makes up human nature, and harsh but entirely convincing criticism is offered of the pretensions of sociobiology. Human behaviour is seen as being on a biological leash, the length and tension of which is not known, which allows humans to create vastly different cultures.

Anyone writing a textbook is confronted with the choice of judiciously summarizing the views of others or of offering an interpretation of their own. Trigg's book convinces me that the latter course is the best. It is possible to respond to his arguments when a fairly clear skeleton lies behind all his description. This makes his attack on relativism exciting and engaging. He does not offer particularly striking arguments to support scientific rationality; but then his fears, given the fabulous triumphs of science, are anyway probably somewhat exaggerated and might have been avoided had he realised that subtle defences of positivism as a cognitive ethic have long been available. He is, however, entirely convincing when dismissing relativism in social science. Primitive peoples are rational in the sense of trying to make sense of and to control their world, but they are not blessed with the *Weltanschauung* of rational science and therefore often arrive at erroneous conclusions: rationality, in other words, is in one sense a human universal and in another the property of some particular people in history. Trigg is right to make this distinction, and to castigate those who wish to consider all belief systems as good as each other: that they are not is amply demonstrated in his eyes by the speed with which cognitively low-powered systems of thought are abandoned once superior systems become available.

The book loses drive and becomes frankly confusing at those points where Trigg does not, as in the case of relativism, have some sort of devil at his back. His concern with human nature seems to drive him towards something like methodological individualism, and this interpretation gains force from critical comments against determinism. Nevertheless, he appreciates the force of social constraint, and makes much of altruism and trust—both of which he feels need social rather than psychologistic, sociobiological or even rational man type explanations. I am unsure even now where he stands in the holism/agent debate. Somewhat similarly, his remorseless occupying of some middle ground as far as human nature is concerned does not carry much force. He notes at one point that it may be that human evolution has depended upon suppressing genetic inheritance, as Freud and Norbert Elias maintained. This seems to me highly likely, and such as to largely discredit sociobiology completely. But given that Trigg takes sociobiology seriously, it would have been helpful to have risked a little more. One has the impression, however, that his conception of human nature is not that far removed from Benthamite man, and that he has no real appreciation of the power of religion.

On the last page of the book, Trigg makes this claim:

> Empirical social science must start from a properly articulated philosophical base if it is to be successful. The philosophy of social sciences cannot be an optional activity for those reluctant to get on with the 'real' empirical work. It is the indispensable starting point for all social science.

Whilst having no wish to make the philosophy of social science a mere optional extra, I am nevertheless *extremely* suspicious of this claim. Philosophy has more often diverted than aided social science, as when it encouraged an entirely meretricious distinction between idealism and materialism which for long misled social scientists. Whilst it is pleasant to see Trigg resolve problems concerning the proper way to understand belief in primitive societies, it is worth remembering that those problems were created by philosophers, notably by Peter Winch, and that the resolution arrived at results from paying close attention to what anthropologists really do. I suspect that the philosophy of social science would advance if it systematically paid greater attention to what is involved in such explanations as social scientists have achieved rather than tell them what to do, particularly as social scientists, including the greatest, have often misinterpreted the nature of their own methodologies. Such attention would lead to a lessening of interest in the debate between holism and agency for social science has *for a very long time* realised that the individual is a social construct. Philosophers should try to codify and theorise good practice, and those concerned with social science would be better advised to reflect upon an achieved major substantive work such as the first volume of Michael Mann's *Sources of Social Power*[1] than to issue directives. Although it is a heresy to say so, I am quite sure that a student learns most philosophy of social science in actual undertaking research, and the moral of this should not be lost on philosophers of social science themselves.

DR. J. A. HALL
University of Southampton and Harvard University

---

[1] M. Mann, *Sources of Social Power. Volume One: From the beginning to 1760 AD*. Cambridge and New York: Cambridge University Press, 1986.

Philosophy and Phenomenological Research

AN INTERNATIONAL QUARTERLY FOUNDED BY MARVIN FARBER

ERNEST SOSA, Editor
RODERICK M. CHISHOLM, Associate Editor

in cooperation with a distinguished group of American and foreign scholars

PPR publishes articles in a wide range of areas including philosophy of mind, epistemology, ethics, metaphysics, and philosophical history of philosophy. No specific methodology or philosophical orientation is required in submissions. An abstract not exceeding 250 words and two nonreturnable copies of any manuscript submitted for publication are required.

Address All Communications to:
Philosophy and Phenomenological Research
BROWN UNIVERSITY, BOX 1947
PROVIDENCE, RHODE ISLAND 02912

Annual subscription rates. $17.00 (foreign, $18.50) for libraries and institutions $13.00 for individuals; single copies $4.25 (double issue, $8.50) and $3.25 (double issue, $6.50), respectively. There is an additional charge for postage and handling on all back issues. A circular listing the main contents of the journal and a list of available back issues will be sent upon request.

THE PHILOSOPHICAL QUARTERLY

Editorial Chairman: Neil Cooper

Founded by the Scots Philosophical Club and the University of St Andrews in 1950, **The Philosophical Quarterly** aims to foster and publish significant contributions in every branch of the subject, promoting discussion of recent philosophical works, and providing critical reviews of recent books, in a style accessible to as wide a group of philosophers as possible.

Published January, April, July, October for the Scots Philosophical Club and the University of St Andrews
Subscription Rates Volume 37, 1987
Individuals: £11 (UK) £13.75 (overseas) US$26 (N. America+Japan)
Institutions: £27.50 (UK) £36.75 (overseas) US$61.50 (N. America+Japan)
Rates include inland or accelerated surface postage.

Please send orders with payment to:

Iris Taylor
Journals Department
Basil Blackwell
108 Cowley Road
Oxford OX4 1JF
England

or to: **Journals Department**
Basil Blackwell
PO Box 1320
Murray Hill Station
NY 10156
USA

Basil Blackwell · Oxford and New York

INQUIRY
An Interdisciplinary Journal of Philosophy

Editor: Alastair Hannay

Volume 30 (1987), Nos. 1-2

Foundations of Wildlife Protection Attitudes	Eugene C. Hargrove
You Can't Get Something for Nothing: Kierkegaard and Heidegger on How Not to Overcome Nihilism	Hubert L. Dreyfus and Jane Rubin
Meaning, Context, and Control: Convergent Trends and Controversial Issues in Current Social-scientific Research on Human Cognition and Communication	Ragnar Rommetveit
Concepts, Structures, and Meanings	G. R. Gillett
Overdue Analysis of Bourdieu's Theory of Practice	Theodore Richard Schatzki
Power and Formation: New Foundations for a Radical Concept of Power	David West
Two Dogmatists	Charles R. Pigden
Tarski - A Dilemma	Richard C. Jennings
Peacocke's Thoughts (Christopher Peacocke: *Thoughts: An Essay on Content)*	Richard E. Aquila

Subscription price:
Institutions: USD 60.-. Individuals: USD 31.-.

Published quarterly by Norwegian University Press, P.O. Box 2959 Tøyen, 0608 Oslo 6, Norway - Publications Expediting Inc., 200 Meacham Ave., Elmont, NY 11003, USA. Editorial address: Institute of Philosophy, P.O. Box 1024 Blindern, 0315 Oslo 3, Norway.

Erkenntnis

An International Journal of Analytic Philosophy

Editors
Carl G. Hempel, *Princeton University, USA;* **Wolfgang Stegmüller,** *Universität München, FRG,* **Wilhelm K. Essler,** *Universität Frankfurt, FRG*

Erkenntnis is a philosophical journal publishing papers on foundational studies and scientific methodology covering:
'Philosophy of Science' and 'Analytic Philosophy'; philosophy of language, of logic, and of mathematics; foundational problems of physics and of other natural sciences; foundations of normative disciplines such as ethics, philosophy of law and of aesthetics; methodology of the social sciences and the humanities; history of scientific method.
Erkenntnis was first published in 1930, and succeeded *Annalen der Philosophie*, which appeared from 1919–29.

Subscription Information **ISSN** 0165–0106
1987, Volume 26–27 (6 issues)
Institutional rate: Dfl. 444.00/US$186.00 incl. postage/handling
Private rate: Dfl. 164.00/US$68.00 incl. postage/handling

Private subscriptions should be sent direct to the publishers

D. Reidel Publishing Company

Kluwer academic publishers group

P.O. Box 989
3300 AZ Dordrecht
The Netherlands

101 Philip Drive
Norwell,
MA 02061, U.S.A.

Falcon House
Queen Square
Lancaster, LA1 1RN, U.K.

dialectica Vol. 40, 1986

Fasc. 3 Contents Sommaire Inhalt

Alan Weir, *Realism and Behaviourism;* P. Swiggers, *Le fondements définitionnels du réalisme constructionnel de Nelson Goodmann;* Clive Stroud-Drinkwater, *Seeing and Speaking of Rabbits;* Neven Sesardić, *Psychology Without Principle of Charity*

Etudes critiques – Herbert Stachowiak, ed., *Pragmatik*; Christopher Peacocke, *Thoughts: An Essay On Content*; W. Balzer/D. Pearce/H.-J. Schmidt, *Reduction in Science*; Clark Zumbach, *The Transcendent Sciences*; François Russo, *Introduction à l'histoire des techniques*; Uri Zelbstein, *L'Univers des Machines*

Fasc. 4 Contents Sommaire Inhalt

Daniel Laurier, *La langue d'une population: le lien entre la sémantique et la pragmatique;* Denis Vernant, *Quantification substitutionnelle, contextes intensionnels et question d'existence;* Jacques Dubucs, *Réalisme et antimécanisme chez K. Gödel;* Marjorie Grene, *Die Einheit des Menschen: Descartes unter den Scholastikern*

Etudes critiques – J. Quéron, *Pic de la Mirandole; H. Bouasse, Instruments à Vent; The Logic of Discourse and the Logic of Scientific Discovery,* J. Hintikka/ F. Vandamme ed.

Received Books — Livres reçus — Eingegangene Bucher; Table annuelle, vol. 40, 1986 — Contents of Vol. 40 — Inhalt Band 40.

Subscriptions Abonnements Abonnemente	Switzerland	Other countries
Subscription rate per annum (4 issues) Abonnement annuel (4 fascicules) Jahresabonnement (4 Hefte)	65.-SFr.	80.-SFr.

Distribution/Auslieferung

Dialectica, Case postale 1081, 2501 Bienne (Suisse)
F.W. Faxon, Stechert Coordinator, 15 Southwest Park, Westwood/Mass. 02090 USA
B.H. Blackwell Ltd., Broad Street, Oxford, England

OXFORD

Mind *The Journal of the Mind Association*

Editor: Simon Blackburn, Pembroke College, Oxford

Mind is the best known British philosophical journal, and one of the most widely read and respected journals in professional philosophy.

Forthcoming and recent papers include *Paul Churchland* on reductive strategies in cognitive neurophysiology, *Ruth Garrett Millikan* on the metaphysical nature of Putnam's anti-realism, *Jonathan Westphal* on colour, *Michael Pendlebury* on force in semantics and *Ian Hacking* on inverse arguments in cosmology.

ORDER FORM

Payment may be made by cheque or major credit card.

Please send me:

☐ *Mind* Volume 96 UK £16 (Individuals £10) Overseas £19 (£13.50)
☐ a sample copy

Name ....

Address ....

Postcode .... Country ....

Oxford University Press, Walton Street, Oxford OX2 6DP, UK

JOURNALS

Brit. J. Phil. Sci. 38 (1987), 1–18 *Printed in Great Britain*

Non-deductive Logic in Mathematics

by JAMES FRANKLIN

1 *Introduction*
2 *Evidence for the Riemann Hypothesis and other conjectures*
3 *Axiomatisation of non-deductive logic*
4 *The problem of induction in mathematics*

Mathematicians often speak of conjectures as being confirmed by evidence that falls short of proof. For their own conjectures, evidence justifies further work in looking for a proof. Those conjectures of mathematics that have long resisted proof, such as Fermat's Last Theorem and the Riemann Hypothesis, have had to be considered in terms of the evidence for and against them. It is argued here that it is not adequate to describe the relation of evidence to hypothesis as 'subjective', 'heuristic' or 'pragmatic', but that there must be an element of what it is *rational* to believe on the evidence, that is, of non-deductive logic.

1 INTRODUCTION

The occurrence of non-deductive logic, or logical probability, in mathematics is an embarrassment. It is embarrassing to mathematicians, used to regarding deductive logic as the only real logic. It is embarrassing for those statisticians who wish to see probability as solely about random processes or relative frequencies: surely there is nothing probabilistic about the truths of mathematics? It is a problem for philosophers who believe that induction is justified not by logic but by natural laws or the 'uniformity of nature': mathematics is the same no matter how lawless nature may be. It is awkward even for proponents of non-deductive logic. If non-deductive logic deals with logical relations weaker than entailment, how can such relations hold between the necessary truths of mathematics?

Previous work on this topic has therefore been rare. There is one notable exception, the pair of books by the mathematician George Polya on *Mathematics and Plausible Reasoning* [1954]. Despite their excellence, they have been little noticed by mathematicians, and even less by philosophers. Undoubtedly this is largely because of Polya's unfortunate choice of the word 'plausible' in his title—'plausible' has a subjective, psychological ring to it, so that the word is almost equivalent to 'convincing' or 'rhetorically persuasive'. Arguments that happen to persuade, for psychological reasons, are rightly regarded as of little interest in mathematics and philosophy. Polya in fact made it clear, however, that he was not concerned with subjective impressions, but with what degree of belief was *justified* by the evidence ([1954] I p. 68). This will be the point of view argued for here, and there will be frequent reference to Polya's books.

Non-deductive logic deals with the support, short of entailment, that some propositions give to others. If a proposition has already been proved true, there is of course no longer any need to consider non-conclusive evidence for it. Consequently, non-deductive logic will be found in mathematics in those areas where mathematicians consider propositions which are not yet proved. These are of two kinds. First there are those that any working mathematician deals with in his preliminary work before finding the proofs he hopes to publish, or indeed before finding the theorems he hopes to prove. The second kind are the long-standing conjectures which have been written about by many mathematicians, but which have resisted proof.

It is obvious on reflection that a mathematician must use non-deductive logic in the first stages of his work on a problem. Mathematics cannot consist just of conjectures, refutations and proofs. Anyone can generate conjectures, but which ones are worth investigating? Which ones are relevant to the problem at hand? Which can be confirmed or refuted in some easy cases, so that there will be some indication of their truth in a reasonable time? Which might be capable of proof by a method in the mathematician's repertoire? Which might follow from someone else's theorem? Which are unlikely to yield an answer until after the next review of tenure? The mathematician must answer these questions to allocate his time and effort. But not all answers to these questions are equally good. To stay employed as a mathematician, he must answer a proportion of them *well*. But to say that some answers are better than others is to admit that some are, on the evidence he has, more reasonable than others, that is, are rationally better supported by the evidence. This is to accept a rolе for non-deductive logic.

The area where a mathematician must make the finest discriminations of this kind—and where he might, in theory, be guilty of professional negligence if he makes the wrong decisions—is as a supervisor advising a prospective Ph.D. student. It is usual for a student beginning a Ph.D. to choose some general field of mathematics and then to approach an expert in the field as a supervisor. The supervisor then chooses a problem in that field for the student to investigate. In mathematics, more than in any other discipline, the initial choice of problem is the crucial event in the Ph.D.-gathering process. The problem must be

(1) unsolved at present
(2) not being worked on by someone who is likely to solve it soon

but most importantly

(3) tractable, that is, probably solvable, or at least partially solvable, by three years' work at the Ph.D. level.

It is recognised that of the enormous number of unsolved problems that have been or could be thought of, the tractable ones form a small proportion, and that it is difficult to discern which they are. The skill in non-deductive logic required of a supervisor is high. Hence the advice to Ph.D. students not to worry too much about what field or problem to choose, but to concentrate on finding a good supervisor.

It is also clear why it is hard to find Ph.D. problems that are also

(4) interesting.

It is not possible to dismiss these non-deductive techniques as simply 'heuristic' or 'pragmatic' or 'subjective'. Although these are correct descriptions as far as they go, they give no insight into the crucial differences among techniques, namely, that some are more reasonable and consistently more successful than others. 'Successful' can mean 'lucky', but 'consistently successful' cannot. 'If you have a lot of lucky breaks, it isn't just an accident', as Marx said (Chandler [1980], p. 560). Many techniques can be heuristic, in the sense of leading to the discovery of a true result, but we are especially interested in those which give reason to believe the truth has been arrived at, and justify further research. Allocation of effort on attempted proofs may be guided by many factors, which can hence be called 'pragmatic', but those which are likely to lead to a completed proof need to be distinguished from those, such as sheer stubbornness, which are not. Opinions on which approaches are likely to be fruitful in solving some problem may differ, and hence be called 'subjective', but the beginning graduate student is not advised to pit his subjective opinion against the experts'. without good reason. Damon Runyon's observation on horse-racing applies equally to courses of study: 'The race is not always to the swift, nor the battle to the strong, but that's the way you bet' (Fadiman [1955], p. 794). An example where the experts agreed on their opinion, and were eventually proved right, is the classification of finite simple groups, described in the next section.

It is true that similar remarks could also be made about *any* attempt to see rational principles at work in the evaluation of hypotheses, not just those in mathematical research. In scientific investigations, various inductive principles obviously produce results, and are not simply dismissed as pragmatic, heuristic or subjective. Yet it is common to suppose that they are not principles of *logic*, but work because of natural laws (or the principle of causality, or the regularity of nature). This option is not available in the mathematical case. Mathematics is true in all worlds, chaotic or regular; any principles governing the relationship between hypothesis and evidence in mathematics can only be logical.

2 EVIDENCE FOR THE RIEMANN HYPOTHESIS AND OTHER CONJECTURES

In modern mathematics, it is usual to cover up the processes leading to the construction of a proof, when publishing it—naturally enough, since once a result is proved, any non-conclusive evidence that existed before the proof is no longer of interest. This was not always the case. Euler, in the eighteenth century, regularly published conjectures which he could not prove, with his evidence for them. He used, for example, some daring and obviously far from rigorous methods to conclude that the infinite sum

$$1+\frac{1}{4}+\frac{1}{9}+\frac{1}{16}+\frac{1}{25}+\cdots.$$

(where the numbers on the bottom of the fractions are the successive squares of whole numbers) is equal to the *prima facie* unlikely value $\frac{\pi^2}{6}$. Finding that the two expressions agreed to seven decimal places, and that a similar method of argument led to the already proved result

$$1-\frac{1}{3}+\frac{1}{5}-\frac{1}{7}+\frac{1}{9}-\frac{1}{11}+\cdots=\frac{\pi}{4}.$$

Euler concluded, 'For our method, which may appear to some as not reliable enough, a great confirmation comes here to light. Therefore, we shall not doubt at all of the other things which are derived by the same method'. He later proved the result (Polya [1954], I, pp. 18–21). A translation of another of Euler's publications devoted to presenting 'such evidence . . . as might be regarded as almost equivalent to a rigorous demonstration' of a proposition is given as a chapter in Polya's books ([1954], I, pp. 91–8).

Even today, in fact, mathematicians occasionally mention in print the evidence that led to a theorem. Since the introduction of computers, and even more since the recent use of symbolic manipulation software packages, it has become possible to collect large amounts of evidence for certain kinds of conjectures [Andrews [1984], pp. 18 & 24).

At present, it is usual to delay publication until proofs have been found. This rule is broken only in work on those long-standing conjectures of mathematics which are believed to be true, but have so far resisted proof. Three of these are especially famous. To non-mathematicians, the best known are probably the conjectures in number theory called Fermat's Last Theorem and Goldbach's Conjecture. The evidence for these will be mentioned below. The most notable to mathematicians themselves, however, is undoubtedly the Riemann Hypothesis.

Riemann stated in his [1859] that he thought it 'very likely' that

> All the roots of the Riemann zeta function (with certain trivial exceptions) have real part equal to $\frac{1}{2}$.

This is the Riemann Hypothesis. The precise meaning of the terms involved is not very difficult to grasp (account in Edwards [1974]) but for the present purpose it is only necessary to observe that this is a simple universal proposition like 'all ravens are black'. It is also true that the roots of the

Riemann zeta function have a natural order, so that one can speak of 'the first million roots' (there are infinitely many roots). Once it became clear that the Riemann Hypothesis would be very hard to prove, it was natural to look for evidence of its truth (or falsity). The simplest kind of evidence would be ordinary induction: Calculate as many of the roots as possible and see if they all have real part $\frac{1}{2}$. This is in principle straightforward (though in practice computational mathematics is difficult, since one needs to devise subtle algorithms which save as much calculation as possible, so that the results can go as far as possible). Such numerical work was begun by Riemann, and was carried on later with the results below:

Worker	Number of roots found to have real part $\frac{1}{2}$
Gram (1903)	15
Backlund (1914)	79
Hutchinson (1925)	138
Titchmarch (1935/6)	1,041

'Broadly speaking, the computations of Gram, Backlund and Hutchinson contributed substantially to the plausibility of the Riemann Hypothesis, but gave no insight into the question of why it might be true' (Edwards [1974], p. 97).

The next investigations were able to use electronic computers, and the results were:

Lehmer (1956)	25,000
Meller (1958)	35,337
Lehman (1966)	250,000
Rosser, Yohe & Schoenfeld (1968)	3,500,000
Brent (1979)	81,000,001

(Table from Brent *et al.* [1982].)

In recent years, work with the latest computers has proceeded with a collaboration between R. P. Brent, in Canberra, and van de Lune, te Riele and Winter, in Amsterdam. The latest result in print is from the Dutch team [1983]: the first 300,000,001 roots of the Riemann zeta function have real part $\frac{1}{2}$. They promise further results. It is one of the largest inductions in the world.

Besides this simple inductive evidence, there are some other reasons for believing that Riemann's Hypothesis is true (and some reasons for doubting it). In favour, there are:

(1) Hardy proved in 1914 that infinitely many roots of the Riemann zeta function have real part $\frac{1}{2}$ (Edwards [1974], pp. 226–9). This is quite a strong consequence of Riemann's Hypothesis, but is not sufficient to make the

Hypothesis highly probable, since if the Riemann Hypothesis is false, it would not be surprising if the exceptions to it were rare.

(2) Riemann himself showed that the Hypothesis implied the 'prime number theorem', then unproved. This theorem was later proved independently. This is an example of the general non-deductive principle that non-trivial consequences of a proposition support it.

(3) Also in 1914, Bohr and Landau proved a theorem roughly expressible as, 'Almost all the roots have real part very close to $\frac{1}{2}$'. More exactly, 'For any $\delta > 0$, all but an infinitesimal proportion of the roots have real part within δ of $\frac{1}{2}$'. This result 'is to this day the strongest theorem on the location of the roots which substantiates the Riemann hypothesis' (Edwards [1974], p. 193).

(4) Studies in number theory revealed areas in which it was natural to consider zeta functions analogous to Riemann's zeta function. In some famous and difficult work, André Weil [1948] proved that the analogue of Riemann's Hypothesis is true for these zeta functions, and his related conjectures for an even more general class of zeta functions were proved to widesprerad applause in the 1970's. 'It seems that they provide some of the best reasons for believing that the Riemann hypothesis is true—for believing, in other words, that there is a profound and as yet uncomprehended number-theoretic phenomenon, one facet of which is that the roots ρ all lie on Re $s = \frac{1}{2}$' (Edwards [1974], p. 298).

(5) Finally, there is the remarkable 'Denjoy's probabilistic interpretation of the Riemann hypothesis' (Edwards [1974], pp. 268–9). If a coin is tossed n times, then of course we expect about $\frac{1}{2}n$ heads and $\frac{1}{2}n$ tails. But we do not expect *exactly* half of each. We can ask, then, what the average deviation from equality is. The answer, as was known by the time of Bernoulli, is $\sqrt{n}$. One exact expression of this fact is:

For any $\varepsilon > 0$, with probability one the number of heads minus the number of tails in n tosses grows less rapidly than $n^{1/2+\varepsilon}$.
(Recall that $n^{1/2}$ is another notation for $\sqrt{n}$.)

Now we form a sequence of 'heads' and 'tails' by the following rule: Go along the sequence of numbers and look at their prime factors. If a number has two or more prime factors equal (*i.e.*, is divisible by a square), do nothing. If not, its prime factors must be all different; if it has an even number of prime factors, write 'heads'. If it has an odd number of prime factors, write 'tails'. The sequence begins:

2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	...
		2^2		2×3		2^3	3^2	2×5		$2^2\times3$		2×7	3×5	2^4		
T	T		T	H	T			H	T		T	H	H		T	...

The resulting sequence is of course not 'random' in the sense of 'probabilistic', since it is totally determined. But it does look 'random' in the sense of 'patternless' or 'erratic' (such sequences are common in number theory, and are studied by the branch of the subject called, perhaps misleadingly, 'probabilistic number theory'; see Kubilius [1964]). From the above, it is likely that

> For any $\varepsilon > 0$, the number of heads minus the number of tails in the first n "tosses" in this sequence grows less rapidly than $n^{1/2+\varepsilon}$.

This statement is equivalent to Riemann's Hypothesis. Edwards comments, in his book on the Riemann zeta function,

> One of the things which makes the Riemann hypothesis so difficult is the fact that there is no plausibility argument, no hint of a reason, however unrigorous, why it should be true. This fact gives some importance to Denjoy's probabilistic interpretation of the Riemann hypothesis which, though it is quite absurd when considered carefully, gives a fleeting glimmer of plausibility to the Riemann hypothesis (Edwards [1974], p. 268).

In the balance against the Riemann Hypothesis there are the following arguments:

(1) Riemann's paper is only a summary of his researches, and he gives no reasons for his belief that the Hypothesis is 'very likely'. No reasons have been found in his unpublished papers. Edwards does give an account, however, of facts which Riemann knew, which would naturally have seemed to him evidence of the Hypothesis. But the facts in question are true only of the early roots; there are some exceptions among the later ones. Edwards concludes:

> The discoveries . . . completely vitiate any argument based on the Riemann–Siegel formula and suggest that, unless some basic cause is operating which has eluded mathematicians for 110 years, occasional roots ρ off the line [*i.e.*, with real part not $\frac{1}{2}$] are altogether possible. In short, although Riemamnn's insight was stupendous it was not supernatural, and what seemed 'probable' to him in 1859 might seem less so today (Edwards [1974], p. 166).

This is an example of the non-deductive rule given by Polya, 'Our confidence in a conjecture can only diminish when a possible ground for the conjecture is exploded' (Polya [1954], II, p. 20).

(2) Although the calculations by computer did not reveal any counterexamples to Riemann's Hypothesis, Lehmer's and later work did unexpectedly find values which it is natural to see as 'near counterexamples'. An extremely close one appeared near the 13,400,000th root (Edwards [1974], pp. 175–9). It is partly this that prompted the calculators to persevere in their labours, since it gave reason to believe that if there were a counterexample, it would probably appear soon. So far, it has not, despite the distance to which computation has proceeded, so Riemann's Hypothesis is not so undermined by this consideration as appeared at first.

(3) Perhaps the most serious reason for doubting Riemann's Hypothesis comes from its connections with the prime number theorem. (The title of Riemann's paper is, 'On the number of primes less than a given magni-

tude'.) This theorem states that the number of primes less than x is (for large x) approximately equal to

$$\int_2^x \frac{dt}{\log t}.$$

If tables are drawn up for the number of primes less than x and the values of

$$\int_2^x \frac{dt}{\log t},$$

for x as far as calculations can reach, then it is always found that the number of primes less than x is actually *less than*

$$\int_2^x \frac{dt}{\log t}.$$

On this evidence, it was thought for many years that this was true for all x. Nevertheless Littlewood proved that this is false. While he did not produce an actual number for which it is false, it appears that the first such number is extremely large—well beyond the range of computer calculations. Edwards comments

> In the light of these observations, the evidence for the Riemann hypothesis provided by the computations of Rosser *et al.* . . . loses all its force (Edwards [1974], p. 269).

This seems too strong a conclusion, since the relevance of Littlewood's discovery to Riemann's Hypothesis is far from clear. But it does give *some* reason to suspect that there may be a very large counterexample to Riemann's Hypothesis, although there are no small ones.

It is plain, then, that there is much more to be said about Riemann's Hypothesis than, 'It is neither proved nor disproved'. Without non-deductive logic, though, nothing more can be said.

The situation with Fermat's Last Theorem and Goldbach's Conjecture is similar, and the evidence for them can be described briefly.

Fermat's Last Theorem, asserted by Fermat about 1637, is

> It is impossible to find whole numbers X, Y, Z and n, with $n \geqslant 3$, such that $X^n + Y^n = Z^n$ (with X, Y, Z not zero) (see Edwards [1977]).

Calculations can be undertaken to check that $X^n + Y^n \neq Z^n$ for small values of X, Y, Z and n. It is often possible, however, for fixed values of the exponent n, to deal with all the (infinitely many) X, Y, Z at once. Fermat proved (with one of the many proofs of number theory that have survived for centuries without any counterexamples, monster-barring, concept-stretching or other approved aids) that there can be no solutions if $n = 4$. Other values were added later, and Kummer in 1847 showed that the Fermat equation has no solutions whenever n is a "regular" prime. The condition for being "regular" is rather complicated; it is believed (but not proved) that most primes are regular, and the first non-regular prime is 37. Later, it was found that if there is a counterexample to Fermat's Last Theorem, then this implies certain conditions on n which can be checked by calculation. The best result in this direction is that of Wagstaff [1978], who checked by computer:

Fermat's Last Theorem is true for all n up to 125,000.

It is also useful to divide Fermat's Last Theorem into two cases according to some rather complicated conditions on X, Y, Z and n. The simpler first case has been checked by computer for n up to 6×10^9 (Lehmer [1981]).

In a letter to Euler of 1742, Goldbach conjectured that

Every even number (except 2) is the sum of two primes.

This conjecture is still neither proved nor disproved. Various consequences of it have been proved (Polya [1954], II, p. 210) and, remarkably, connections have appeared between Goldbach's Conjecture and Riemann's Hypothesis. Hardy and Littlewood proved in 1924 that a generalisation of Riemann's Hypothesis and a certain estimate implied that *most* even integers are the sum of two primes. Vinogradov in [1937] showed that every sufficiently large odd integer is the sum of three primes, and these methods were soon adapted to show Hardy and Littlewood's result without any assumptions. In [1948] Renyi found that every even number is the sum of a prime and an 'almost prime' (a number with few prime factors). Linnik in [1952] showed that Riemann's Hypothesis itself implied a proposition relevant to Goldbach's Conjecture. Results on the problem are still sometimes found, but there do not seem to have been dramatic advances in the last thirty years.

A last mathematical example of the central role of non-deductive inference is provided by the classification of finite simple groups, one of the great cooperative efforts of modern pure mathematics. As a case study, it has the merit that the non-deductive character of certain aspects was admitted rather explicitly by the principals. This happened because of the size of the project. Since so many people were involved, living in different continents and working over some years, it was necessary to present partial findings in print and at conferences, with explanations as to how these bore on the overall results hoped for.

Groups are one of the basic abstract entities of mathematics, having uses in describing symmetry, in classifying the various kinds of curved surfaces and in many other areas. ('Group' is a technical term, and does not just mean 'collection' as it does in English.) To read the following it is only necessary to know:

(1) A group consists of finitely or infinitely many 'members'; the number of members of a finite group is called its 'order'.

(2) Any group is composed, in a certain sense, of 'simple' groups. ('Simple',

> like 'group', is a technical term; 'simple' groups are not in any sense easy to understand, but are so-called because they are not composed of smaller groups.)

A fundamental question is then: how many different finite simple groups are there? And what is the order of each? (One needs of course to prove the answers are correct.) It is these questions that were attacked by the classification of finite groups project.

The project proper covers the twenty years from 1962 to 1981 inclusive. Groups had been studied in the nineteenth and early twentieth centuries, and various finite simple groups were found. It was discovered that most of them fell into a number of infinite families. These families were quite well described by the mid-1950's, with some mopping-up operations later. There were, however, five finite simple groups left over from this classification, called the Mathieu groups after their discoverer in the 1860's. Around 1960, it was not known whether any more should be expected, or, if not, how much work it might take to prove that these were the only possible ones.

The field was opened up by the celebrated theorem of Feit and Thompson (of Cornell and Chicago Universities respectively) in [1963]:

> The order of any finite simple group is an even number.

Though the result is easy to state and understand, their proof required an entire 255-page issue of the *Pacific Journal of Mathematics*. This theorem is a consequence of the full classification result (since if one knew all the finite simple groups, one could simply check that the order of each of them was even). It thus appeared that if the full classification could be found at all, it would be a vast undertaking.

In fact, the final step in the answer was added in February,1981. The full proof is spread over some 300 to 500 journal papers, taking up somewhere between 5,000 and 10,000 pages (Gorenstein [1982], p. 1).

We wish to examine the logical situation as the proof developed, noting particularly the confidence—justified as it happened—that the workers in the field had in the answer long before the end was reached.

It turned out that the five Mathieu groups were not the only 'sporadic' groups, as groups outside the infinite families came to be called. The first new one was discovered by Zvonimir Janko [1966], in Canberra, and excitement ran high as researchers applied many methods and discovered more. The final tally of sporadic groups stands at 26. These 'discoveries' had in many cases a strong non-deductive aspect, as explained by Daniel Gorenstein, of Rutgers, who later became the father-figure of the project and leading expert on how it was progressing:

> Another aspect of sporadic group theory makes the analogy with elementary particle theory even more apt. In a number of cases (primarily but not exclusively those in which computer calculations were ultimately required) "discovery" did not include

> the actual construction of a group—all that was established was strong evidence for the existence of a simple group G satisfying some specified set of conditions X. The operative metamathematical principle is this: if the investigation of an arbitrary group G having property X does not lead to a contradiction but rather to a "compatible" internal subgroup structure, then there exists an actual group with property X. In all cases, the principle has been vindicated; however, the interval between discovery and construction has varied from a few months to several years (Gorenstein [1982], pp. 3–4).

Michael Aschbacher, another leader in the field in the 1970's, distinguished three stages for any new group: discovery, existence and uniqueness.

> I understand a sporadic group to be discovered when a sufficient amount of self consistent information about the group is available . . . Notice that under this definition the group can be discovered before it is shown to exist . . . Of course the group is said to exist when there is a proof that there exists some finite simple group satisfying P . . . (Aschbacher [1980], pp. 6–7).

Some groups attracted more suspicion than others; for example that discovered by Richard Lyons was for some time habitually denoted Ly? and spoken of in such terms as, 'If this group exists, it has the following properties . . .' (Tits [1969/70], p. 204). Lyons entitled his original paper 'Evidence for the existence of a new finite simple group'. A similar situation arose with another of the later groups, discovered by O'Nan. His paper, 'Some evidence for the existence of a new simple group', was devoted to finding 'some properties of the new simple group G, whose existence is pointed at by the above theorems' (O'Nan [1976], p. 422).

The rate of discovery of new sporadic groups slowed down after 1970, and attention turned more to the problem of showing there were no more possible. At a conference at the University of Chicago in 1972 Gorenstein laid out a sixteen-point program for completing the classification (Gorenstein [1979]). It was thought over-optimistic at the time, but immense strides were soon made by Aschbacher, Glauberman and others, more or less following Gorenstein's program.

> The turning point undoubtedly occurred at the 1976 summer conference in Duluth, Minnesota. The theorems presented there were so strong that the audience was unable to avoid the conclusion that the full classification could not be far off. From that point on, the practicing finite group theorists became increasingly convinced that the "end was near"—at first within five years, then within two years, and finally momentarily. Residual skepticism was confined largely to the general mathematical community, which quite reasonably would not accept at face value the assertion that the classification theorem was "almost proved" (Gorenstein [1982], pp. 5–6).

(Notice that 'almost proved' indeed does not mean anything in deductive logic. With hindsight, one can say that a theorem was almost proved when most of the steps in the proof were found; but before a proof is complete, there can only be good reason to believe that a sequence of existing steps will constitute most of a future proof).

By the time of the conference at Durham, England in 1978 (described in its Proceedings as on 'the classification of simple groups, a programme which is now almost complete') optimism ran even higher. At that stage existence and uniqueness had been proved for 24 of the sporadic groups, leaving two 'for which considerable evidence exists' (Collins [1980], p. 21). One of these was successfully dealt with in 1980 ('four years after Janko's initial evidence for such a sporadic group' (Gorenstein [1982], p. 110)), and attention focussed on the last one, known as the 'Monster' because of its immense size (order about 10^{54}).

> That the search for sporadic groups was not totally haphazard can be seen from the remarkable simultaneous realization by Fischer in West Germany and Griess in the United States in 1974 that there might be a simple group having a covering group . . . (Gorenstein [1982], p. 92).

Consequences of the existence of this group were then studied:

> Soon after the initial "discovery", Griess, Conway and Norton noticed that every nontrivial irreducible character of a group G of type F_1 has degree at least 196,883 and very likely such a group G must have a character of this exact degree. Indeed, on this assumption, Fischer, D. Livingstone and Thorne eventually computed the full character table of such a group G (Gorenstein [1982], pp. 126–7).

Aschbacher, lecturing at Yale in 1978, said ([1980], p. 13)

> When the Monster was discovered it was observed that, if the group existed, it must contain two new sporadic groups (the groups denoted by F_3 and F_5 in Table 2) whose existence had not been suspected up to that time. That is, these groups were discovered as subgroups of the Monster. Since that time the groups F_3 and F_5 have been shown to exist. This is analogous to the situation in the physical sciences where a theory is constructed which predicts certain physical phenomena that are later verified experimentally. Such verification is usually interpreted as evidence that the theory is correct. In this case, I take the existence of F_3 and F_5 to be very good evidence that the Monster exists.

He added, concerning the prospects of discovering more sporadic groups:

> My belief is that there are at most a few groups yet to be discovered. If I were to bet, I would say no more (pp. 14–15).

Gorenstein's survey article of 1978 (Gorenstein [1979]) contains perhaps the experts' last sop to deductivism, the thesis that all logic is deductive. He wrote:

> At the present time the determination of all finite simple groups is very nearly complete. Such an assertion is obviously presumptuous, if not meaningless, since one does not speak of theorems as "almost proved" (pp. 50–1).

To the deductivist, the fact that most steps in a proposed proof are completed is no reason to believe that the rest will be.

Undeterred, however, Gorenstein went on to say:

> The complete proof, when it is obtained, will run to well over 5,000 journal pages! Moreover, it is likely that at the present time more than 80% of these pages exist (p. 51)

and

> The assertion that the classification is nearly complete is really a prediction that the presently available techniques will be sufficient to deal with the problems still outstanding. In its support, we cite the fact that, with two exceptions, all open questions are open because no one has yet examined them and not because they involve some intrinsic difficulty (p. 52).

A year after the Durham conference, the experts assembled again at Santa Cruz, California, in a mood of what could only be called supreme confidence. Gorenstein's survey opened with the remark:

> My aim here is to present a brief outline of the classification of the finite simple groups, now rapidly nearing completion (Gorenstein [1980], p. 3).

Another contributor to the conference began his talk:

> Now that the problem of classifying finite simple groups is probably close to completion . . . (Hunt [1980], p. 507).

What concern there was was less about the completion of the project than about what to do next; the editor of the conference proceedings began by commenting, 'In the last year or so there have been widespread rumors that group theory is finished, that there is nothing more to be done' (Mason [1980], p,. xiii). These became something more than rumours when the *New York Times* Week in Review (June 22, 1980) headlined an article 'A School of Theorists Works Itself Out of a Job'.

All this confidence proved justified. Griess was able to show the existence of the Monster, and finally, in 1981, Simon Norton of Cambridge University completed the proof of the uniqueness of the Monster (Gorenstein [1982], p. 1).

3 AXIOMATISATION OF NON-DEDUCTIVE LOGIC

The correctness of the above arguments is not affected by the success or failure of any attempts to formalise, or give axioms for, the notion of non-deductive support between propositions. Many fields of study, such as geometry in the time of Pythagoras or pattern-recognition today, have yielded bodies of truths while still resisting reduction to formal rules. Even so, it is natural to ask whether the concept *is* easily formalisable. This is not the place for detailed discussion, since the problem has nothing to do with mathematics, and has been dealt with mainly in the context of the philosophy of science. The axiomatisation that has proved serviceable is the familiar axiom system of conditional probability: if h (for 'hypothesis') and e (for 'evidence') are two propositions, $\mathrm{P}(h,e)$ is a number between 0 and 1 inclusive expressing the degree to which h is supported by e, which satisfies

$$\mathrm{P}(\text{not-}h,e) = 1-\mathrm{P}(h,e)$$

$$\mathrm{P}(h',h\,\&\,e)\times\mathrm{P}(h,e) = \mathrm{P}(h,h'\,\&\,e)\times\mathrm{P}(h',e)$$

While some authors (Carnap [1950], Stove [1973]) have been satisfied with this system, others (Keynes [1921], Koopman [1940]) have thought it too strong to attribute an exact number to P(*h*,*e*) in all cases, and have weakened the axioms accordingly. Their modifications are essentially minor.

Needless to say, command of these principles alone will not make anyone a shrewd judge of hypotheses, any more than perfection in deductive logic will make him a great mathematician. To achieve fame in mathematics, it is only necessary to string together enough deductive steps to prove an interesting proposition, and submit the results to *Inventiones Mathematicae*. The trick is finding the steps. Similarly in non-deductive logic, the problem is not in knowing the principles, but in bringing to bear the relevant evidence.

The principles nevertheless provide *some* help in deciding what evidence will be helpful in confirming the truth of a hypothesis. It is easy to derive from the above axioms the principle

If h & b implies e, but $P(e,b) < 1$, then $P(h,e \,\&\, b) > P(h,b)$.

If *h* is thought of as hypothesis, *b* as background information, and *e* as new evidence, this principle can be expressed as 'The verification of a consequence renders a conjecture more probable', in Polya's words ([1954], II, p. 5). He calls this the 'fundamental inductive pattern'; its use was amply illustrated in the examples in section 2 above. Further patterns of inductive inference, with mathematical examples, are given in Polya ([1954), ch. XII–XIII).

There is one point that needs to be made precise especially in applying these rules in mathematics. If *e* entails *h*, then P(*h*,*e*) is 1. But in mathematics, the typical case is that *e* does entail *h*, though this is perhaps as yet unknown. If, however, P(*h*,*e*) is really 1, how is it possible in the meantime to discuss the (non-deductive) support that *e* may give to *h*, that is, to treat P(*h*,*e*) as not equal to 1? In other words, if *h* and *e* are necessarily true or false, how can P(*h*,*e*) be other than 0 or 1?

The answer is that, in both deductive and non-deductive logic, there can be *many* logical relations between two propositions; some may be known and some not. To take an artificially simple example in deductive logic, consider the argument

If all men are mortal, then this man is mortal
All men are mortal

Therefore, this man is mortal

The premises entail the conclusion, certainly, but there is more to it than that. They entail the conclusion in two ways: firstly, by *modus ponens*, and secondly by instantiation from the second premise alone. More complicated and realistic cases are common in the mathematical literature. Feit and

Thompson's proof that all finite simple groups have even order, occupying 255 pages, was simplified by Bender [1970]. That means that Bender found a different, and shorter, logical route from the definition of 'finite simple group' to the proposition, 'All finite simple groups have even order' than the one known to Feit and Thompson.

Now just as there can be two deductive paths between premises and conclusion, so there can be a deductive and non-deductive path, with only the latter known. Before the Greeks' development of deductive geometry, it was possible to argue

All equilateral (plane) triangles so far measured have been found to be equiangular
This triangle is equilateral

Therefore, this triangle is equiangular

There is a non-deductive logical relation between the premises and the conclusion; the premises support the conclusion. But when deductive geometry appeared, it was found that there was also a deductive relation, since the second premise alone entails the conclusion. This discovery in no way vitiates the correctness of the previous non-deductive reasoning, or casts doubt on the existence of the non-deductive relation.

4 THE PROBLEM OF INDUCTION IN MATHEMATICS

That non-deductive logic is used in mathematics is important first of all to mathematics. But there is also some wider significance for philosophy, in relation to the problem of induction, or inference from the observed to the unobserved.

It is common to discuss induction using only examples from the natural world, such as, 'All observed flames have been hot, so the next flame observed will be hot' and 'All observed ravens have been black, so the next observed raven will be black'. This has encouraged the view that the problem of induction should be solved in terms of natural laws (or causes, or dispositions, or the regularity of nature) that provide a kind of cement to bind the observed to the unobserved. The difficulty for such a view is that it does not apply to mathematics, where induction works just as well as in natural science.

Examples of this were given above in section 2, in connection with Riemann's Hypothesis, but let us take a particularly straightforward case:

$$\frac{\text{The first million digits of } \pi \text{ are random}}{\text{Therefore, the second million digits of } \pi \text{ are random.}}$$

('Random' here means 'without pattern', not 'probabilistically generated'.)

The number π has the decimal expansion

3·14159265358979323846264338327950288419716939937 . . .

There is no apparent pattern in these numbers. The first million digits have in fact been written down (Guilloud and Bouyer [1974]). Inspection of these digits reveals no pattern, and computer calculations can confirm this impression. It can then be argued inductively that the second million digits will likewise exhibit no pattern. This induction is a good one (indeed, everyone believes that the digits of π continue to be random indefinitely, though there is no proof), and there seems to be no reason to distinguish the reasoning involved here from that used in inductions about flames or ravens. But the digits of π are the same in all possible worlds, whatever natural laws may hold in them, or fail to. Any reasoning about π is also rational or otherwise, regardless of any empirical facts about natural laws. Therefore, induction can be rational independently of whether there are natural laws.

This argument does not show that natural laws have no place in discussing induction. It may be that mathematical examples of induction are rational because there are *mathematical* laws, and that the aim with scientific examples is to find some substitute, such as natural laws, which will take the place of mathematical laws in accounting for the continuance of regularity. But if this line of reasoning is pursued, it is clear that simply making the supposition, 'There are laws', is of little help in making inductive inferences. No doubt mathematics is completely lawlike, but this does not help at all in deciding whether the digits of π continue to be random. In the absence of any proofs, induction is needed to support the law (if it is a law), 'The digits of π are random', rather than the law giving support to the induction. Either 'The digits of π are random' or 'The digits of π are not random' is a law, but in the absence of knowledge as to which, we are left only with the confirmation the evidence gives to the first of these hypotheses. Thus consideration of a mathematical example reveals what can be lost sight of in the search for laws: laws or no laws, non-deductive logic is needed to make inductive inferences.

It is worth noting that there are also mathematical analogues of Goodman's 'grue' paradox. Let a number be called 'prue' if its decimal expansion is random for the first million digits and 6 thereafter. The predicate 'prue' is like 'grue' in not being projectible. 'π is random for the first million digits' is logically equivalent to 'π is prue for the first million digits', but this proposition supports 'π is random always', not 'π is prue'. Any solutions to the 'grue' paradox must allow projectible or 'natural' properties to be found not only in nature but also in mathematics.

These considerations illustrate Polya's remark that non-deductive logic is better appreciated in mathematics than in the natural sciences ([1954], II, p. 24). In mathematics there can be no confusion over natural laws, the regularity of nature, approximations, propensities, the theory-ladenness of observation, pragmatics, scientific revolutions, the social relations of science or any other red herrings. There are only the hypothesis, the evidence and the logical relations between them.

University of New South Wales

REFERENCES

ANDREWS, G. E. [1984]: 'On the Wall Polymonials and the L-M-W conjectures', *J. Australian Math. Soc.*, Ser. A, **37**, pp. 17–26.

ASCHBACHER, M. [1980]: *The Finite Simple Groups and Their Classification*. Yale University Press.

BENDER, H. [1970]: 'On the Uniqueness Theorem', *Illinois J. Math.*, **14**, pp. 376–84.

BRENT, R. P., VAN DE LUNE, J., TE RIELE, H. J. J. and WINTER, D. T. [1982]: 'On the Zeros of the Riemann Zeta Function in the Critical Strip (II)', *Mathematics of Computation*, **39**, pp. 681–88.

CARNAP, R. [1950]: *Logical Foundations of Probability*. Routledge.

CHANDLER, C. [1980]: *Hello, I Must Be Going: Groucho Marx and His Friends*. Sphere.

COLLINS, M. J. (*ed.*) [1980]: *Finite Simple Groups II*. Academic

COOPERSTEIN, B. and MASON, G. (*eds.*) [1980]: *The Santa Cruz Conference on Finite Groups, Proc. of Symposia in Pure Mathematics*. Amer. Math. Soc., vol. **37**.

DAVIS, P. J. and HERSH, R. [1981]: *The Mathematical Experience*. Birkhauser. Ch. 8.

EDWARDS, H. M. [1974]: *Riemann's Zeta Function*. Academic.

EDWARDS, H. M. [1977]: *Fermat's Last Theorem*. Springer.

FADIMAN, C. [1955]: *The American Treasury*. Harper.

FEIT, W. and THOMPSON, J. [1963]: 'Solvability of Groups of Odd Order', *Pacific J. Math.*, **13**, pp. 775–1029.

GORENSTEIN, D. [1979]: 'The Classification of Finite Simple Groups (I)', *Bull. Amer. Math. Soc.* New Series, **1**, pp. 43–199.

GORENSTEIN, D. [1980]: 'An Outline of the Classification of Finite Simple Groups', in COOPERSTEIN and MASON [1980], pp. 3–28.

GORENSTEIN, D. [1982]: *Finite Simple Groups*. Plenum.

GUILLOUD, J. and BOUYER, M. [1974]: *1,000,000 de decimales de π*. Commissariat à l'energie Atomique.

HUNT, D. C. [1980]: 'A Computer-based Atlas of Finite Simple Groups', in COOPERSTEIN and MASON [1980], pp. 507–10.

JAMES, R. D. [1949]: 'Recent Progress in the Goldbach Problem', *Bull. Amer. Math Soc.*, **55**, pp. 246–60.

JANKO, Z. [1966]: 'A New Finite Simple Group With Abelian 2-Sylow Subgroups and Its Characterization', *J. Algebra*, **3**, pp. 147–86.

KEYNES, J. M. [1921]: *A Treatise on Probability*. Macmillan.

KOLATA, G. B. [1976]: 'Mathematical Proofs: The Genesis of Reasonable Doubt', *Science*, **192**, pp. 989–90.

KOOPMAN, B. O. [1940]: 'The Axioms and Algebra of Intuitive Probability', *Annals of Mathematics*, **41**, pp. 269–92.

KUBILIUS, J. [1964]: *Probabilistic Methods in the Theory of Numbers*. Amer. Math. Soc. Translations of Mathematical Monographs, no. 11.

LEHMER, D. H. [1981]: 'On Fermat's Quotient, Base Two', *Mathematics of Computation*, **36**, pp. 289–90.

LINNIK, YU. [1952]: 'Some Conditional Theorems Concerning the Binary Goldbach Problem', *Izv. Akad. Nauk SSSR*, **16**, pp. 503–20.

LYONS, R. [1972]: 'Evidence for a New Finite Simple Group', *J. Algebra*, **20**, p. 540–69.

MASON, G. [1980]: Preface to COOPERSTEIN and MASON (1980).

O'NAN, M. [1976]: 'Some Evidence for the Existence of a New Finite Simple Group', *Proc. London Math. Soc.*, **32**, pp. 421–79.

POLYA, G. [1954]: *Mathematics and Plausible Reasoning* (vol. I, *Induction and Analogy in Mathematics*, and vol. II, *Patterns of Plausible Inference*). Princeton University Press.

RENYI, A. [1962]: 'On the Representation of an Even Number as the Sum of a Prime and an Almost Prime', *Amer. Math. Soc. Translations* (2), **19**, pp. 299–321.

RIEMANN, B. [1859]: 'On the Number of Primes Less Than a Given Magnitude', translated in EDWARDS [1974], pp. 299–305.

STOVE, D: C. [1973]. *Probability and Hume's Inductive Scepticism*. Oxford University Press.

TITS, J. [1969/70]: 'Groupes Finis Simples Sporadiques', *Séminaire Bourbaki*, in *Springer Lecture Notes* no. 180.

VAN DE LUNE, J. and TE RIELE, H. J. J. [1983]: 'On the Zeros of the Riemann Zeta Function in the Critical Strip (III)', *Mathematics of Computation*, **41**, pp. 759–67.

WAGSTAFF, S. S. [1978]: 'The Irregular Primes to 125000', ***Mathematics of Computation*, 32**, pp. 583–91.

WEIL, A. [1948]: *Courbes Algébriques et Variétés Abéliennes*. Hermann.

Brit. J. Phil. Sci. 38 (1987), 19–26 *Printed in Great Britain*

Realism and Antirealism in Artificial Intelligence[1]

by DAVID H. HELMAN

1 *Introduction*
2 *Two Analyses of the Propositional Attitudes*
3 *Realism and Antirealism in Artificial Intelligence*
4 *The Reconciliation of Realism and Antirealism*

In the philosophy of mind, the controversy between realists and antirealists often concerns the logical form of sentences embedded in attitude reports. Antirealists believe that such sentences refer to psychological states; realists believe that they refer to situations or states of the world. In this essay, it is shown how these two modes of semantic representation are associated with different approaches to the computational modeling of cognitive processes. I put forward a normative account of methodology in artificial intelligence that reconciles realism and antirealism, and suggest that my account sheds light on the problem of representing the logical form of attitude reports.

1 INTRODUCTION

In the philosophy of mind, the controversy between realists and antirealists generally concerns the semantics or logical form of sentences embedded in reports of propositional attitudes. An antirealist believes that the interpretation or reference of a sentence embedded in an attitude report is a psychological state of the agent who is the subject of the attitude report (See Bencivenga [1983]). An antirealist, for example, might believe that 'It is raining' refers to a truth-value when that sentence stands on its own, but refers in an attitude report, such as 'John believes it is raining', to John's image, idea, or conception of rain. For a realist, the interpretation or reference of a sentence is a state of the world and not a state of mind, whether or not the sentence is embedded in an attitude report (see Barwise and Perry [1983]).

---

[1] Part of this research was carried out while I was a Visiting Scholar at the Stanford Center for the Study of Language and Information. My research at the Stanford Center for the Study of Language and Information was partially funded by the Case Western Reserve Center Automation and Intelligent Systems Research. My thanks to both centers for their assistance. I am also indebted to Dr. Ivan Blair for his comments on an earlier draft of this essay. The views and conclusions expressed in this document are those of the author and should not be interpreted as representative of the views of the Stanford Center for the Study of Language and Information or the Case Western Reserve Center for Automation and Intelligent Systems Research.

In this essay, I would like to show how these two styles of semantic representation, the realist mode and the antirealist mode, are associated with different theories of mental representation in artificial intelligence. First, I will argue that the realist and antirealist views in semantics imply, respectively, realist and antirealist research programs in the computational modeling of cognitive states. I will then put forward a normative account of methodology in artificial intelligence that reconciles the realist and antirealist approaches towards the computatinal modeling of cognition. Finally, I will suggest that my account of the methodology of artificial intelligence sheds light back on the problem of representing the logical form of attitude reports.

2 TWO ANALYSES OF THE PROPOSITIONAL ATTITUDES

Suppose the realist analyzes:

(1) John believes it is raining.

along the lines of:

(2) There exists a situation *S* and John stands in a certain relation *R* to *S*.

Consider the realist who allows that for John to stand in relation *R* to *S*, it is necessary for John to be in state *P*. Suppose the antirealist analyzes (1) as:

(3) John has a certain psychological state *Q*.

The realist position outlined above and antirealism are the same, psychologically speaking, if states *P* and *Q* are the same. The point is that if two semantic analyses of the attitudes constitute distinct psychological theories, then their analyses of the propositional attitudes must imply distinct accounts of mental representation.

How can an analysis of the semantics of attitude reports imply a theory of mental representation? Let us consider analysis (3) of sentence (1) above. Suppose one thinks that sentences embedded in attitude reports refer to peculiarly mental states. Then one must also think that having an attitude involves standing in one or another relation to a mental state. It is difficult, furthermore, to see how one could believe that having an attitude involves being in particular psychological states, but not believe that the laws of the mind include generalizations that relate these kinds of states. In this way, we may connect antirealism in semantics to antirealism in the theory of mental representations, *i.e.*, the view that fundamental properties of the mind are a consequence of the laws that relate mental contents or representations. The paradigm proponent of this latter view is Hume, who holds that the fundamental properties of the mind are determined by the laws that govern the association of ideas (Hume [1739]).

We may count as antirealists in artificial intelligence researchers who attempt to simulate mental processes by means of computational procedures which mirror tenets of associationist or Humean psychology (Minsky [1981], Schank [1982]). For example, one simple thesis of Humean psychology is that if we have experienced the constant conjunction of A and B, and we now experience A, we will expect B. This kind of expectation provides the fundamental link between distinct representations in a number of computational models of memory (see Abelson and Schank [1977], Wilensky [1983]).

A realist believes that the ontology of a theory of mind need not consist in peculiarly mental objects. Believing, thinking, and knowing, according to the realist, relate agents to states of the world. What does realism imply in the theory of mental representations? Consider an antirealist theory of mind T with an ontology of ideas and structures derived from ideas, and containing associationist laws that relate those ideas and structures. Consider also a realist theory of mind S with an ontology of situations or events. We may have to permute the laws of the association of ideas in theory T only slightly to get laws that relate the events or situations in theory S. For example, theory T might say that whenever A occurs an agent will typically have a certain idea K, and whenever B occurs, an agent will typically have an idea J. If K and J are related by a law of the association of ideas, then, according to theory T, when an agent has K, that agent will usually have J shortly thereafter. Let us consider the realist theory S. According to that theory, K and J are unnecessary intermediaries. Theory S states that an agent can be directly related to the situation A and B. Suppose that theory S also says that whenever an agent is related to the situation A, she will be related to situation B shortly thereafter. That is, suppose the laws of theory S posit the same causal sequences as the laws of theory T, however, in theory S we replace reference to ideas by references to real situations. Theory S and theory T are not, in this case, significantly different theories of mental representation. If there is a significant difference between an antirealist and realist theory of mental representation, this difference cannot merely consist in a difference between the ontologies of the two theories.

It was noted above that if the ontology of a theory of mind consists in peculiarly mental elements, then the laws of the corresponding theory of the mind should be psychological laws. A similar inference can be made with regard to realist theories of the propositional attitudes. If an adequate account of the mind must refer to real-world events or situations, then the laws of the mind, at least in part, must concern the law-like relationships between real-world events or situations. What makes for a realist theory of mental representations, therefore, is that one does not posit uniquely psychological laws to relate objects posited by the theory. Realists represent the mind by representing the world. That is, the connections between distinct mental representations, or between aspects of a single mental representation, are dictated in realist representations of mind by laws which relate to the external world and not by the laws of the association of ideas.

A realist account of mental representations should be founded in laws that concern external objects and events, however, these laws must also take into account aspects of human psychology. The realist, for example, may be interested in answering the following question: What laws of motion must we suppose a person implicitly 'knows' in order to account for his/her ability to grasp objects? Research pertaining to this problem may borrow from psychology as well as physics. Such research is, however, distinguished from Humean psychology by its concern to explicate the knowledge of the subject in the form of commonsense or qualitative laws about the real world, apart from the tendencies of a subject to associate types of ideas or experiences (see Hayes [1982], De Kleer and Brown [1984], Forbus [1984]).

3 REALISM AND ANTIREALISM IN ARTIFICIAL INTELLIGENCE

The realist-antirealist distinction in the theory of mental representations is an important distinction with regard to artificial intelligence research. This point can be illustrated if we consider two ways in which graphs have been used as data structures in artificial intelligence. In a typical antirealist graph representation structure, the nodes of the graph will be atoms or lists of atoms which are supposed to represent individual concepts or ideas. These nodes will be linked by arcs labeled with atoms or lists of atoms that are supposed to represent relationships between mental representations. In the *Politics* program of Carbonell [1981], for example, nodes in directed graphs are supposed to represent the beliefs of an agent about the goals of a political actor (*e.g.*, the Soviet Union). These nodes are linked by arcs representing the relationships between these beliefs. For example, the program of Carbonell may represent an agent as believing that one perceived goal of a political actor is relatively less important than another, or that one perceived goal of a nation is a sub-goal of another.

Realist accounts of cognition in artificial intelligence have also used graphs as data structures. These programs do not, however, attempt to match patterns of preprogrammed 'conceptual' connections. In a realist computational representation, the links between nodes in a graph represent the interactions between parts of a system as they are constrained by causal laws. Consider, for example, the following account of Patil *et al* [1982]:

> To summarize, a causal link in the system is an object denoting the causal relation between a cause-effect pair. It specifies a multivariate relation between various aspects of the cause and the effect, taking into account the context and the assumption under which the causal relation is being instantiated. . . . The mapping relation permits the computation of a value for any of the attributes of the cause or effect

> given some (perhaps incomplete) subset of the values of the other attributes. A common use of this facility is to compute some aspect of the description of an effect from knowing its cause; e.g., the severity of the effect from the severity and duration of the cause. The mapping relation may also be used to infer in the opposite direction, for example permitting the calculation of the duration of the cause given its severity and the severity of its effects (p. 201).

In Patil's realist analysis of medical knowledge, for a doctor to know the relationship between dehydration and water-loss is for the doctor to be able to calculate, given a context and certain assumptions about the physiology of the patient, how water-loss partially determines dehydration. In this style of representation, notions such as expectation and belief seem to be dispensable. They need not figure in the representation language of the program. And there need not be any sort of straightforward mapping from the types of causal inferences that are made by the system to types of propositions referring to mental events.

Realist representation structures appear to be far superior to antirealist representation structures. Suppose we try to represent dehydration as a partial function of water-loss in an antirealist theory. It seems that the best we could do is have explicit links in a graph for each type of situation we anticipate, that is, the program could represent explicitly the knowledge that if water loss is between X and Y, and the context is C, and the background assumptions are B, then we may expect dehydration between D and E. Notice that there is a great deal of information lost in this style of representation—we have in effect collapsed all values of water loss between X and Y. We can, though, in a realist representation scheme, represent the continuously varying values for water loss numerically, semi-quantitatively (see Widman [1986]), or qualitatively (note Kuipers [1985]).

Antirealist representation structures are also relatively inflexible. Suppose that the patient suffering from dehydration is a child, so the background assumptions regarding patient physiology are quite different. In an antirealist representation scheme, we must write a completely new graph to capture the different expectations that derive from these changed background assumptions, for we will have to adjust all the conditions on the links of the graph to adjust for the new background conditions. But graphs are individuated by just these conditions. Therefore, on the account of the antirealist, the graph which captures the relationship between water loss and dehydration in adults and the graph that captures the relationship between water loss and dehydration in children will be as different as an apple and an orange. If the laws of the realist theory are good enough, however, it will capture the close relationship between dehydration in adults and dehydration in children. That is, one set of functions may be able to compute the necessary values for dehydration given the quite different description of water loss in the child and the adult.

When relationships are considered that are, physically or physiologically speaking, well understood, the research program of the realist in artificial intelligence looks considerably more promising than the research program of the antirealist. Antirealist representation methods, however, are not designed to be used in systems that represent low-level physical or physiological regularities. Scripts (Abelson and Schank [1977]), for example, are supposed to represent our expectations regarding typical sequences of events. In a typical restaurant sequence, we might order food, eat the food and pay. A script-based program, therefore, might explicitly link the event of ordering the food and the event of eating food. There is, however, no procedure known, that when given information about the physical environment of a restaurant during an ordering of food, and information regarding the physiological conditions of the employees of the restaurant and the patrons at that time, will derive the information that food will be served. We clearly do not understand the relationship between ordering food and receiving it as a consequence of our implicit knowledge of such a procedure.

4 THE RECONCILIATION OF REALISM AND ANTIREALISM

One might think, given the above discussion, that realist representations of cognitive processes will be adequate as long as we represent very simple situations; situations in which a physical description of the interrelationships between parts of the situation are computationally tractable. I am not inclined to dismiss this idea, but it can be misleading. Consider the following situations:

(1) I have one copy of a manuscript which I have labored on for several years. My daughter puts it in the oven. I see the manuscript in the oven, and scream 'The *manuscript* is in the oven'.

(2) I have many copies of a manuscript which I have been working on for several years. My daughter puts one of these copies in the oven, and knowing that the oven is likely to be damaged if the manuscript ignites, I scream 'The manuscript is in the *oven*' (these examples follow Dretske [1977]).

The physical descriptions of the situations at the time of my screaming can be the same in event one and event two. Therefore, even though these descriptions might be relatively simple, they cannot adequately represent my cognitive state in both cases, for I am prepared to reason and to act quite differently in the two situations. The realist might reply that the situation at the time of my screaming in example 1 may be differentiated from the situation at the time of my screaming in example 2 if we take a wider perspective, *i.e.*, if we take into account the causal history that relates these cases to past events. My cry in example 1 is causally determined by the events in which I have labored over my manuscript. My cry in example 2 is tied causally to my experiences involving the oven.

The reply of the realist, however satisfactory it may or may not be from a metaphysical point of view, is completely inadequate from a cognitive perspective. Suppose it is true that my causal history completely determines my present mental state; that my causal history completely determines whether I am concerned with the manuscript or the oven. It does not follow that the mental representations which allow me to reason and react appropriately in these situations are representations of that causal history. That is, my cognitive states may be brought about as a consequence of my life history, without these cognitive states being a representation of my life history. It is unlikely, furthermore, that I could represent all the events causally related to the present one; there are simply too many events so related. It seems that the representations that differentiate the case where I am concerned with the manuscript from the case where I am concerned with the oven can only be *partial* representations of my history (see Barwise and Perry [1983]). The events in such a partial history need not be causally related to one another, in any scientifically acceptable sense of 'cause'. For example, when I walk into a restaurant, I receive a menu. We may say, perhaps, that my walking into the restaurant 'caused' me to receive a menu. As noted above, however, it would be difficult to give a physical account of the relationship between my walking into the restaurant and my receiving a menu. It is doubtful that this is a relationship that would emerge as a consequence of the laws of a qualitative or commonsense physics.

The above reflections lead to a reconciliation between the realist and antirealist perspectives in artificial intelligence. The realist perspective is appropriate when we represent a limited number of factors that may be related functionally (if the past history of the situation plays a causal role, this must be accounted for functionally). Antirealist representations are appropriate when we need to note which of our past experiences are relevant in the present case (and where this relevance relation cannot, because of our limited resources, be represented functionally). The recording of our past experiences enables us to call up routines or ways of acting that apply to present exigencies. Once we have decided how we will react in a situation, our ability to represent the world realistically enables us to be flexible. For example, once I decide I must push the bookcase (because, based upon my past experience, this is the appropriate way to act), lower-level feedback loops, based on functional representations of the relationships between the relevant physical parameters, will adjust my actions to the bookcase at hand (Helman [1986]).

One way to sum up the conclusions of this essay is that there must be at least two levels of representation in any successful simulation of cognition, a level that partially represents aspects of the life history of the organism, and a level that functionally or realistically represents the environment of the organism (see Pylyshyn [1981], Beer *et al* [1985]). This conclusion, if it is correct, gives us an interesting new perspective on the realism-antirealism controversy in the philosophy of mind. Do sentences embedded in attitude reports refer to psychological states or do they refer to states of the world?

Brit. J. Phil. Sci. 38 (1987), 27–34 *Printed in Great Britain*

Teleology and Logical Form

by ANDREW KERNOHAN

Recent proposals by Taylor, Bennett, Wright and Cohen to identify teleological systems as systems governed by teleological laws and teleological laws as laws of a certain logical form are discussed. Suggested logical forms are treated with both extensional and simple non-extensional models of nomic necessity and shown to generate problematic entailments not derivable from the causal form alone.

Teleology had, for many years, a bad press. Its defenders faced problems such as explaining how an organism's goals, which are future states of affairs, could function as final causes of its present behaviour. The particular promise of teleology is a means of grounding psychological explanations that otherwise would be left appealing to the beliefs and intentions of organisms. But performing this service in a satisfactory and non-question-begging manner requires the formulation of a non-intentional teleology where talk of purposes and goals is made innocuous. Recently, though, teleology has returned to some favour with solutions being offered for these problems and more. What is common to such proposals are the suggestions that, first, teleological systems are systems which are governed by teleological *laws* and, second, teleological laws are just like ordinary laws of nature except that they have a different logical *form*. Together these suggestions, if they can be made to work, would remove the major mysteries surrounding teleology. In this paper I want to examine some difficulties with the second part of this programme. The trouble with putting a lot of weight on logical form to ground a distinction is that one's formulations sometimes turn out to be logically equivalent to formulae which have much less plausibility. I will show that this is what happens to several suggestions about the logical form of teleological laws. This is not a knock-down argument against the possible ultimate success of the programme since we cannot hope to canvass all conceivable formulations, but the considerations can be seen either as casting some doubt on the programme's viability or as indicating an area for further research.

Ordinary laws of nature have, in the simplest possible case, the following logical form:

(1) For any system x, x is a Q because x is a P.

In this formula, it is natural to see x being P as the cause of x being Q. To put this into logical notation we would have to arrive at some understanding of the connective 'because' and its alternatives: 'leads to', 'produces', 'causes', etc. But the problem of choosing a connective also confronts the regimentation of teleological laws, so we can put it aside for the moment. The essential thing is that in comparing the logical form of teleological and causal laws we do so relative to the same choice of connective. In its simplest form, a teleological law looks like this:

(2) For any system x, x does B because x is such that x doing B leads to x getting G.

I take it that the talk of doings and gettings in (2) does not irrevocably distinguish it from (1). If we want to use the copula we can write, though clumsily, 'x is exhibiting-a-doing-B' or 'x is exhibiting-a-getting-G'. Note also that the quantification in (2) is universal and not just over a restricted class of teleological systems. The programme being discussed proposes to identify teleological systems as ones which obey teleological laws. So a prior restriction of the domain of teleological laws would beg the question in hand. What does distinguish a teleological from a causal law is the characteristic way in which the antecedent state of affairs is described. Instead of being described by a simple monadic predicate as in the open sentence, 'x is a P', the description used by a teleological law has a characteristic forward looking structure, 'x is such that x doing B leads to x getting G'. If this distinction can be formalized and shown to be logically legitimate, we would have a way of distinguishing teleological laws.

Furthermore, if the strategy worked it would avoid the two main *prima facie* problems facing teleology. It is not a psychological or intentional description of a system to say that it is such that its doing B leads to its getting G. The letter 'G' was picked here to point up the analogy between G and a goal, but unless some intentional operator sneaks into this characterization of the antecedent state of the system, it is not really a goal of the system. And, though the system is described in terms of future consequences, it is still described *now* in such terms. x will do B in the future because it is now such that its doing B will lead to its getting G.

Before attempting to make trouble for this attractive proposal, let me show that it is not a creature of my own invention by citing some of the relevant literature. The origins of the proposal go back to Charles Taylor's book *The Explanation of Behaviour*. There he writes:

> To offer a teleological explanation of some event or class of events, e.g., the behaviour of some being, is, then, to account for it by laws in terms of which an event's occurring is held to be dependent on that event's being required for some end. To say that the behaviour of a given system should be explained in terms of purpose, then, is, in fact, to make an assertion about the form of the laws, or the type of laws which hold of the system. . . . The condition of an event B occurring is, then, not a certain state of P [purpose], but that the state of the system S and the environment E be such that B is required for the end G, by which the system's purpose is defined. ([1964], p. 9)

This passage clearly indicates the two step process whereby teleological purpose is first reduced to subsumption by teleological laws and then these laws characterized by their form. The next passage indicates how Taylor thinks this strategy of logical form can cure the psychological affliction of teleology.

> To say that a system can only be explained in terms of purpose, then—at least insofar as this is an assertion about the form of the laws—does not involve making an unverifiable claim any more than it involves postulating an unobservable entity. The element of 'purposiveness' in a given system, the inherent tendency towards a certain end, which is conveyed by saying that the events happen 'for the sake of' the end, cannot be identified as a special entity which directs the behaviour from within, but consists rather in the fact that in beings with a purpose an event's being required for a given end is a sufficient condition of its occurrence. ([1964], p. 10)

It is worth noting that if we interpret B's being required for G as B's being a necessary condition for G and put this together with what is said about sufficient conditions in the last passage, the following formalization is suggested for teleological laws.

$$(3)\quad (x)\{(Gx \rightarrow Bx) \rightarrow Bx\}$$

This formulation is of course subject to the usual sorts of counterexamples to reading causation in terms of material implication as well as some new ones of its own. But let it stand for the moment.

Noting this last point, Wright explicitly resists commitment to any particular connective for 'because' or 'tends to bring about'. He gives what he calls the 'modified Taylor formula' for the form of a teleological explanation.

> S does B for the sake of G iff:
> (i) B tends to bring about G
> (ii) B occurs because (i.e. is brought about by the fact that) it tends to bring about G. ([1976], p. 39)

This has the form of a deductive-nomological explanation. (i) is a statement of the initial conditions and (ii) is the teleological law. Notice the adherence to Taylor's strategy. But if we are going to assess properly the use of logical form we are going to have to ignore Wright's caginess about the nature of the connective.

Footnoting Taylor, Jonathan Bennett gives the following logical form for a teleological law. He writes,

> As a start, I count as a teleological law any lawlike statement of the form.
> $(x)(f)(t)((Rx \text{ at } t \;\&\; f/Gx \text{ at } t) \rightarrow fx \text{ at } t+d)$ . . . ([1976], p. 39)

The notation 'f/Gx' is used for what he terms 'instrumental predicates'. These are earlier defined as follows:

> F/Ga at time T if and only if at T a is so structured and situated that the truth of Fa at a later time is causally required for and causally sufficient for the truth of Ga at a still later time. (p. 38)

Later in his exposition (p. 71) Bennett drops the 'causally required for' part of his definition. By his use of the terms 'causality' and 'lawlike' Bennett obviously intends a teleological law to be assembled out of more than just logically necessary and/or sufficient conditions. But if we again ignore this for a moment, and ignore the quantification over time (t) and properties (f), and also ignore the readiness condition (Rx), spelling out the instrumental predicate a law would look like

$$(4)\quad (x)\{(Fx \leftrightarrow Gx) \rightarrow Fx\}$$

or in the later revision

$$(5)\quad (x)\{Fx \rightarrow Gx) \rightarrow Fx\}$$

As a final example of this strategy of identifying teleological laws as laws of a particular kind consider this one from the philosophy of history. G. A. Cohen introduces the notion of what he calls a consequence law:

> A consequence law is a universal conditional statement whose antecedent is a hypothetical causal statement. A consequent law relevant to the explanation of an event (as opposed to an object's having a certain property) takes this form:
>
> IF it is the case that if an event of type E were to occur at t_1, then it would bring about an event of type F at t_2, THEN an event of type E occurs at t_3. (Cohen, [1979], pp. 259–260)

He also gives a schema for laws which support explanations of objects having properties.

> IF it is true of an object o that if it were F at t_1, then it would, as a result, be E at t_2, THEN o is F at t_3. ([1979], p. 261)

In a footnote he gives the following formal version:

$$(x)\{(Fx \text{ at } t_1 \,\Box\!\!\rightarrow Gx \text{ at } t_2) \longrightarrow (Fx \text{ at } t_3)\} \text{ (p. 261)}$$

On the troublesome matter of the nature of the connectives involved, Cohen writes that the boxed arrow is 'a connective expressing hypothetical causation, and the unboxed arrow is to be interpreted in whatever is the right way of interpreting the arrow between the antecedent and the consequent of a statement of natural law.' ([1979] n. p. 261)

It might be wondered if these authors perhaps did not have something other than logical form in mind when they talked of 'forms' or 'kinds' of laws. But it is not clear what else this could have been. The versions quoted are schemas for laws, not laws themselves; the *BG*'s and *EF*'s function as predicate variables and have no content in themselves on which a distinction of kinds of laws could be made. So if the distinction between causal and teleological law schemas cannot be made on content, all that remains is logical form.

To properly assess this characterization of teleology and to check out its logical legitimacy, we have to fill in details of the causal connective.

Wright disagrees when, after criticizing Taylor for committing himself to a necessary and/or sufficient conditions analysis of causality, he writes:

> Analysis has to stop somewhere, of course; and when we get something as objectively demonstrable as is a causal connection in a vast array of useful cases, we should at least have the perspicacity to stop there: We are not likely to find anything better. Certainly this counsel is preferable to going down with an attractive over-simplification under a hail of counterexamples. ([1976], pp. 33–34)

This would be a fair position on the matter if we could be sure that teleology would only inherit the problems of analyses of causality *and no new ones of its own*. But we cannot be sure of this, and we cannot find out unless we look at particular proposals for the logical form of teleological laws which include a commitment on the nature of the causal connective and see if we find any new problems that an analogous causal law would not have to face. Such new problems need not show themselves until formalization is attempted.

As one might expect, forms for teleological laws like (3), (4) and (5) are not promising. The trouble is that they can be shown to entail (because logically equivalent) forms which are totally implausible. These entailments are not entailments of the corresponding causal law, so they are new problems for these purported teleological forms. Taking the causal connective as giving sufficient conditions, it is straightforward to show that a teleological law of the form '$(x)\{(Bx \rightarrow Gx) \rightarrow Bx\}$' is logically equivalent to a law of the form '$(x)Bx$'. Here is a proof. Moving right to left we have for any arbitrary individual a, Ba. A true conclusion is implied by any antecedent so we have that $(Ba \rightarrow Ga) \rightarrow Ba$. Since the individual is arbitrary, $(x)\{(Bx \rightarrow Gx) \rightarrow Bx)$. Moving left to right, assume for purposes of reductio that not $(x)Bx$, i.e. that there is something that is not Bx. So there is some individual c such that $\sim Bc$. But the left hand side, instantiated by c, implies $(Bc \rightarrow Gc) \rightarrow Bc$. By modus tollens we have $\sim(Bc \rightarrow Gc)$, that is $Bc \,\&\, \sim Gc$. Simplifying we get Bc which contradicts our assumption. So the equivalence holds. This is, of course, not acceptable. We do not want a teleological law to say simply that it is a law that everything does B. Nor can the problem be avoided by changing the connective used. If we go to necessary and sufficient conditions, it can be shown that the formula '$(x)\{(Bx \leftrightarrow Gx) \leftrightarrow Bx\}$' is logically equivalent to '$(x)Gx$', which is just as bad. If we go to necessary conditions, teleological laws turn out to be tautologies. Taylor's favourite form, (3), and Bennett's first try, (4), can both be shown to be logically equivalent to the formula '$(x)(Bx \vee Gx)$' and hence to each other. But it is implausible that everything is either doing B or getting G; my coffee mug, for one, certainly is not.

Predictably these extensional characterizations of teleological form do not work. Worse, they generate new problematic entailments not generated by simple causal laws alone. The obvious next thing to do is to try a non-extensional characterization. Scientific laws differ from accidentally true universal generalizations in having a certain necessity. This is not full-blown logical necessity—truth in all possible worlds—but a more restricted version. Let us define a new sentence operator as follows: '$\underline{\mathit{Law}}A$' is true if, and only if, 'A' is true at all the nomically possible worlds. Analogously, define '$\underline{\mathit{Pos}}A$' to be true if, and only if, 'A' is true at some nomically possible world. This will enable us to say something about the difference in modal force between laws and accidental generalizations. The account is, of course, incomplete until something is said about the determination of membership in the set of nomically possible worlds. But notice that this is a common problem for accounts of both causal and teleological laws, so we can safely set it aside (making only the assumption that there is at least one other nomically possible world) and look for additional problems that are generated by the peculiar logical form of teleological laws.

Let us interpret causality as giving nomically necessary sufficient conditions to get the following suggestion for the logical form of a teleological law:

(6) $\underline{\mathit{Law}}(x)\{\underline{\mathit{Law}}(Bx \rightarrow Gx) \rightarrow Bx\}$

What is particularly attractive about this form is that it no longer permits derivation of the unacceptable conclusions we got with the extensional formulation. From (6) it is not possible to deduce that everything does B or that it is a law that everything does B. To prove this, we have to produce a model in which (6) is true but '$(x)Bx$' (and hence '$\underline{\mathit{Law}}(x)Bx$') is false. Such a model is one with two nomically possible worlds and one individual, a, with 'Ba' false and 'Ga' true at the actual world and 'Ba' true and 'Ga' false at the second world. This model makes the conclusion, 'Ba', false at our world and also falsifies '$\underline{\mathit{Law}}Ba$', so we just need to show that it makes true '$\underline{\mathit{Law}}\{\underline{\mathit{Law}}(Ba \rightarrow Ga) \rightarrow Ba\}$'. Since '$Ba$' is false at the actual world, the antecedent, '$\underline{\mathit{Law}}(Ba \rightarrow Ga)$' must be chosen false as well. To falsify '$\underline{\mathit{Law}}(Ba \rightarrow Ga)$' we must falsify '$Ba \rightarrow Ga$' at some nomically possible world. This cannot be done at the actual world, but it is done at the second world where 'Ba' is true and 'Ga' is false. So '$\underline{\mathit{Law}}(Ba \rightarrow Ga)$' is false at both worlds, and thus '$\underline{\mathit{Law}}(Ba \rightarrow Ga) \rightarrow Ba)$' is true at both by falsity of antecedent. Therefore '$\underline{\mathit{Law}}\{\underline{\mathit{Law}}(Ba \rightarrow Ga) \rightarrow Ba\}$' is true, which is what needed to be proved. Similar results hold for other interpretations of the connective besides material implication. This formulation avoids the pitfalls of the extensional interpretation in a rather nice way.

But interesting as this result is, there are problems with accepting it. The problems do not arise, as one might think at first glance, from any sort of use/mention equivocation. (6) seems to say, in a causal interpretation of nomicity, that a law that $Bx \rightarrow Gx$ causes Bx. But how can a law, which is, after all, a linguistic item, cause anything? Now this reading would be particularly tempting if the scope of the quantifiers in (6) were different, so that we had instead,

(7) $\underline{Law}(x)\{\underline{Law}(y)(By \rightarrow Gy) \rightarrow Bx\}$

It is definitely not clear from (7) what is doing the causing. But (6) is clear about the cause; it is whatever instantiates x and has the property of being such that it is a law that $Bx \rightarrow Gx$. So we can avoid this objection if we are not tempted by the modification of (6) shown in (7).

The problem with (6) is that the following false sentence can be derived from it:

(8) $(x)\underline{Pos}Bx$

Before saying what I think is wrong with (8), let me give a quick proof that it follows from (6). Suppose that it does not. Then there will be something, c, such that it is not nomically possible that Bc. That is it will be true that $\underline{Law} \sim Bc$. Assuming that everything nomically necessary is true, we get $\sim Bc$. Appropriately instantiating (6) for the actual world and using modus tollens, it follows that $\sim \underline{Law}(Bc \rightarrow Gc)$. This implies $\underline{Pos}(Bc \;\&\; \sim Gc)$. If a conjunction is nomically possible, then each of the conjuncts must be, so $\underline{Pos}Bc$. But this contradicts our original assumption that $\sim \underline{Pos}Bc$. So (8) does follow from (6). (8) says that every actually existent thing is such that it is nomically possible that it does B. But this cannot be true because teleological laws are required only to explain systems which exhibit considerable complexity of behaviour (B). (8) says that everything, including simple things like electrons, must be capable of this behaviour. It is highly implausible that there is a nomically possible world where electrons exhibit the complex behaviour of even a simple teleological system such as an amoeba, let alone a person. Notice that (8) does not say that $\underline{Pos}(x)Bx$. It seems unlikely, but there may be a nomically possible world where simple things like electrons do not exist and everything is complex enough to do B. However, (8) quantifies over everything in our actual world. Nor can we get round this consequence by restricting the application of (6), and thus (8), to teleological systems. Teleological systems are identified as systems which obey teleological laws. Teleological laws are supposedly identified by their logical form. Since this claim is what is in question, it would be circular to assume we could identify teleological systems in advance of being able to characterize the distinctive form of teleological laws. We cannot identify teleological systems without first identifying teleological laws.

Another problem that emerges as peculiar to teleological laws is their iteration of the law operator. (6) requires sentences of the form '$\underline{Law}\underline{Law}A$' as well as ones of the form '$\underline{Law}A$'. But how can we make sense of the notion that it is a law that it is a law that A? Let us try a scientific realist approach. Suppose that it is a law that P leads to Q. Then the law is made true by virtue of some real features of P and Q. Now further suppose that we have reached that golden age yet to come when science has finished its job and the world is completely known. All the laws of physics, biology, psychology etc. have been discovered. At that point *all* the real features of the world will have been described; all the natural kinds which back the laws of science will have been found. If the laws are made true by real causal features of the world, there will be no real features left over to account for the truth of its being a law that it is a law that *P* leads to *Q*. If the truth of '*LawLawA*' is accounted for by the same real features of the world that account for the truth of '*LawA*', then we cannot make separate sense of iterated law operators.

Neither of the two proposals for teleological form which we have examined have worked. Of course this does not show that no possible proposal can work, though the systematic way in which teleological forms generate new problems not afflicting simple nomological forms may be some evidence that way. I have tried playing around with formulations either involving the logic of counterfactuals or incorporating temporal variables but without much success. However this may be due to lack of wit on my part and not the intrinsic insolubility of the problem. My moral, then, is that if this strategy for salvaging teleology is to be successful, research must be done to find a formulation which embodies a plausible logic of causality without creating new problems which the causal logic alone does not face.

Dalhousie University, Halifax, Nova Scotia, Canada

REFERENCES

Bennett, J. [1976]: *Linguistic Behaviour*. Cambridge University Press.
Cohen, G. A. [1979]: *Karl Marx's Theory of History*. Princeton University Press.
Taylor, C. [1964]: *The Explanation of Behaviour*. Routledge and Kegan Paul.
Wright, L. [1976]: *Teleological Explanations*. University of California Press.

Brit. J. Phil. Sci. 38 (1987), 35–47 *Printed in Great Britain*

Theories of Rationality and Principles of Charity

by ROBERT WACHBROIT

While there is no consensus over how the "principle of charity" should be formulated, there is nevertheless a general agreement that the use of that principle is required in attributions of rationality. What has perhaps not been generally noted is that taking the principle of charity seriously results in a conception of rationality that is at odds with the conception underlying many discussions concerning rationality, especially in the philosophy of science. What I would like to do here is articulate this conception. My purpose is not to produce a refutation of the standard conception. Rather, it is to challenge the thought that unless the standard conception is correct, the result is skepticism. By challenging this thought, the way is open for some non-skeptical doubts.

A theory of scientific rationality falls into two parts, depending upon whether we are dealing with the rationality of scientific actions or of scientific beliefs. One part is concerned with the various activities characteristic of science—choosing research projects, designing and conducting experiments, etc. An account of rationality here consists in providing a framework for identifying a goal or aim—sometimes gravely referred to as "the aim of science"—and showing that there is an appropriate connection between the activity in question and the goal. The other part is concerned with scientific theories. An account of rationality here consists in providing a framework for judging whether the reasons for believing these theories are good reasons.[1] Although I will confine myself to this part of a theory of rationality, I hope it will be clear how the following remarks are to be modified to apply to the other part.

The traditional approach to articulating a theory of rationality proceeds by formulating criteria or rules specifying the conditions under which one thing is a good reason for believing something else. The standard work in confirmation theory takes its problem to be that of explicating the conditions under which the qualitative or quantitative confirmation-relation holds.[1] Thus, Carnap makes it quite plain that the success of his project requires the formulation of explicit rules of confirmation. And Hempel argues that the objectivity of scientific validation requires that there be a rational reconstruction of it in terms of rules.[2] Other philosophers, who may not be as concerned to develop a precise theory of confirmation but who still claim that science develops in a rational manner, take their task to be that of formulating the rules according to which the rational development of science proceeds. Thus, Newton-Smith claims that anyone who holds that science is rational is committed to articulating a set of principles for assessing theories. Lakatos asserts that a theory of scientific rationality consists of a set of rules for the appraisal of theories. Laudan claims that as far as actual scientific disputes go, they are rationally terminated by appeal to shared rules of rationality.[3] Even Feyerabend, whose attacks on the rationality of science are notorious, assumes a criteria-based conception of rationality.[4] How else could an attack on method be an attack on rationality?

---

[1] These are *two* parts. Talk of the goals or aims of a belief seems to be completely out of place. One cannot choose to believe something in order to realize some goal in the way that one can choose to do something in order to realize some goal. (Otherwise, self-deception would be as easy as looking the other way.) Thus, Laudan ([1984], esp. chapter 3) overstates the case by suggesting that rationality in all cases involves identifying a goal. See also Newton-Smith [1980], p. 4.

While it should be noted that some of these claims are qualified by remarks such as that there is no algorithm for rational theory choice and that the rules for rationality are not mechanical, it is not at all clear what these qualifications come to. Newton-Smith argues that the set of rules has to be supplemented by a reference to a faculty of judgment, where this is understood as a skill on the model of wine-connoisseurship (233ff). Nevertheless, when a rationale is given for appealing to this skill—its rationale is based on its success (270)—it becomes clear that this appeal to judgment is no more than an appeal to a reliable technique whose inner workings we may not (yet) entirely understand. No real restriction on the ability of rules to capture rationality is being acknowledged, rather a point about evidence is being made: *i.e.*, that the judgments of scientists should be considered as part of the evidence for a theory. Lakatos's position is harder to characterize since his enterprise is to articulate a research program rather than a theory of rationality. Thus, even though at any stage the proposed rules will not capture all of the accepted, rational judgments, this need not reflect badly on the research program (131ff). Nevertheless, this situation drives the program so as to propose rules that will encompass more of the particular judgments not previously captured. Hence, the ideal appears to be that of rationality as completely captured by rules.

The criteria approach has thus been the dominant approach even though there has not been much in the way of acknowledged success in identifying rules of rationality. The project of formulating rules has certainly not been straightforward, for it has encountered formal problems, such as Goodman's "new riddle of induction", as well as problems from the history of science, such as alleged conflicts between what some set of rules enjoin and observed scientific practice.[1] What supports the traditional approach must presumably be some general argument. However, the traditional approach is often not explicitly stated, much less argued for. When it is argued for, the arguments appear to be one of two types.

---

[1] I am assuming throughout that there is a fairly straightforward connection between *Q* being evidence for *P* and *Q* being a reason for believing that *P* (or, if the enterprise is more skeptical, accepting that *P*—see note 1, p. 39), such as that articulated in Achinstein [1978], p. 27.

[2] See Carnap [1962], Section 2; and Hempel, "Studies in the Logic of Confirmation", originally published in 1945, reprinted in Hempel [1965].

[3] See Newton-Smith [1980], p. 9; Lakatos [1978], p. 103; and Laudan [1984], p. 32.

[4] See, for example, Feyerabend [1975], p. 155.

One type of argument turns on an appeal to objectivity. According to Hempel's version of this argument, either there are rules governing the rational acceptability of theories or their acceptability is a matter of a "sense of evidence" or a "feeling of plausibility".[2] But the latter can be excluded since that is a confusion of logical and psychological (i.e., subjective) considerations. The objectivity of rationality thus requires the existence of criteria of rationality.

The other type of argument turns on the claim that a theory of rationality is a (theoretical) reconstruction of (correct) reasoning. If Q is a good reason for believing that P, then it must be possible for us to reason from Q to P. But reasoning from one thought to another is not a random process. Nor is it merely any lawlike sequence of thoughts, beginning with the thought that Q and ending with the thought that P. Reasoning is an activity guided by certain rules, *i.e.*, the canons of reasoning. Accordingly, in reasoning from Q to P we are following certain rules, which define the reason-relation. Therefore, unless we deny that there is reasoning, there must be rules of reasoning and, hence, criteria of rationality.

In short, on the traditional and still widely accepted approach, a theory of rationality consists in formulating rules of rationality. In so far as no such set of rules can be formulated or in so far as actual scientific practice does not conform to such rules, then, to that extent, a conclusion of skepticism, subjectivism, or irrationalism should be drawn.

Nevertheless, the alternative arising from taking the principle of charity seriously indicates that this conclusion does not follow. In order to show this, let me begin by offering a general characterization of the principle of charity that can locate conveniently the points of disagreements in the different versions of that principle. For these purposes, I will divide the principle of charity into two claims: (i) ascribing rationality to others requires discerning a similarity between them and us; (ii) discerning this similarity proceeds by imaginatively projecting ourselves into their situation.

(i) There have been considerable disagreements over what sort of similarity the principle of charity requires. Davidson's position is the most well-known. According to him, discerning a similarity in the case of charity consists in discerning a general agreement over what is true. Thus, charity in attributions of rationality requires that most of the others' beliefs are true, by our lights.[1] Some writers have objected to this by giving examples of communities or individuals who rationally hold to beliefs which are false by our lights.[2] However, it is impossible to assess this dispute without a means for counting beliefs and identifying the totality of an individual's beliefs. It is always open to Davidson to reply that such critics ignore the widespread agreement, much of it too obvious to bother about mentioning, which forms the backdrop to these striking and dramatic cases of disagreement. Thus, as it stands, Davidson's version of the similarity claim is quite vague. A different version of the similarity claim proceeds by arguing that simply maximizing the number of true beliefs, by our lights, need not yield the most reasonable account since it does not pay sufficient attention to the speakers' reasonable errors, by our lights. The discerned similarity should be a similarity in the relationships that obtain between their beliefs, desires, perceptions, available evidence, *etc.*, and ours.[3] On this view, the rationality of a belief does not depend on simply the content of the belief.[4]

---

[1] For a discussion of some further problems, see Hempel [1981].

[2] Hempel [1965], p. 9.

(ii) However the requisite similarity is understood, the similarity is not held to be one between the others, there and then, and us, here and now. Rather, the similarity is to be between the others, there and then, and us, *were* we there and then. For example, although it would be irrational for us, here and now, to believe in geocentric astronomy, it certainly wasn't irrational for ancient astronomers to believe in geocentric astronomy, nor would it be irrational for us to believe in geocentric astronomy were we in their epistemic position. This is plain because what is rational to believe depends upon what evidence is available. The principle of charity reflects this in its second claim, that discerning the requisite similarity proceeds via imaginatively projecting ourselves into the others' situation. While there have not been as many disagreements over this second claim as there have been over the first, those that have taken place turn, for the most part, on whether discerning the requisite similarity *must* proceed via this imaginative projection. That is to say, although many acknowledge that we do rely upon an imaginative projection to discern the requisite similarity, there is some disagreement over whether this reliance is necessary[5] or it is merely a stand in for a set of criteria to be discovered by some future science of human nature.[6]

This last disagreement can be seen as a disagreement over how seriously to take the principle of charity. By taking the principle of charity seriously

---

[1] See, for example, Davidson [1984], p. 27.

[2] See, for example, Thagard and Nisbett [1983].

[3] See, for example, Grandy [1973] and McGinn [1977].

[4] Arguably, this view might capture whatever intuitions underlie Davidson's view since a similarity in structure combined with a similarity in circumstances would tend to result in similarities in the contents of beliefs.

[5] Harman [1982], p. 571.

[6] Lewis [1983], p. 112.

I mean taking this second claim to be necessary: the only way of discerning the similarity requisite for an attribution of rationality is by imaginatively projecting ourselves into the others' situation. There are no other kinds of criteria. This yields a different conception of rationality. Let me call a theory of rationality based on this conception, *i.e.*, one that takes the principle of charity seriously in this way, a "charity-based theory of rationality".

What is the alternative that a charity-based theory of rationality presents? In order to better describe this alternative, I should state clearly what I take to be the point of constructing a theory of rationality.

Without attempting a complete typology, there are at least three tasks or aims in constructing a theory of rationality, distinguishable though not always thought worth distinguishing. One task is to form a reply to skeptical doubts about our claims to rationality. A theory of rationality so understood is a theory of justification or a theory of knowledge as traditionally conceived in epistemology. In order to achieve this, the theory is argued to have authority to assess our current judgments, even though, in practice, any widespread conflict is more cause for revising our theory of rationality than for revising wholesale our current judgments. The work in confirmation or probability theory that takes seriously the traditional problem of induction exemplifies this task.[1] Criteria of evidence, and hence of rationality, are formulated which presumably underwrite most of the current scientific theories even though they are argued for on other grounds. Indeed, the difference between confirmation theory and epistemology seems to be entirely in the sophistication of examples—a difference between electrons and tomatoes.

A different task for a theory of rationality is that of providing an analysis of what rational judgments consists in. This may be a matter of formulating a definition of good reasons, evidence, or related concepts or it may be a matter of providing a more robust account that incorporates facts of psychology and sociology. In either case—whether it be conceptual analysis or its natural[2] offspring—the significance of this task hangs on the terms of the analysis. Can rationality be understood entirely in terms of deductive logic or must we appeal to considerations concerning probability, explanation, or our actual scientific tradition? A good deal of the work in confirmation theory since Hempel's classic paper[3] exemplifies this aim.

A third task for a theory of rationality is that of articulating a standpoint from which we can adjudicate controversial attributions of rationality. This task is often pursued in the context of articulating how the history of science can be properly written as a history of rational inquiry. How much of science can be written as intellectual, internal history and how much can only be written as social, external history? In order to achieve this the theory judges whether the reasons given by the proponents (or opponents) of some scientific change—a change in theory or a change in research activity—were actually good reasons. If they were, then the history of that change can be written as a piece of internal history. If not, then we can only write an external history of that change, looking to other factors—hidden agendas, socioeconomic conditions, etc.—to make sense of what happened then.[1] The point of this third task does not rest on any assumptions about whether the approaches of internal and external history are conflicting or complementary. A theory of rationality is needed to determine the limits of internal history, even though the limits of external history may be quite different.

---

[1] A good example of this is Reichenbach [1938]. A concern with replying to skepticism can take many forms, including that of conceding some of the skeptic's doubts. Thus, although van Fraassen has not yet presented an epistemology to go along with his constructive empiricism, as outlined in his [1980], it is clear that he will try to show that observational data or, more generally, the virtues of a theory (cf. his [1983]) provide reasons for acceptance rather than belief, where acceptance, in his sense, is less vulnerable to skeptical doubts than belief.

[2] The question of its legitimacy is of course another matter.

[3] Hempel, *op. cit.*

These tasks are by no means competing. Indeed, a number of approaches discussed in the literature can be seen as aiming at more than one of these tasks since, by construing the underlying motivations in a certain way, satisfying one task arguably satisfies another. For example, the first task—that of satisfying the skeptic—and the third task—that of settling attributions of rationality—need not be separated if the third task is also seen to be motivated by skeptical concerns. There seems to be little difference, aside from scope and emphasis, between asking whether any (scientific) judgment is rational and asking whether any of the judgments of past scientists were rational. If we had criteria of rationality meeting the requirements of the first task, accomplishing the third task would be simply a matter of applying those criteria to interesting episodes of the past.

Nevertheless, skepticism need no more underlie the settling of attributions of rationality than it need underlie the study of the history of science. There are many reasons for studying the history of science having nothing to do with finding an answer to skepticism. And yet, any history of science that is more than a sociological account of some influential, human institution must appeal to a theory of rationality which indicates where an intellectual history can be written, when the reasons announced were reasons enough.[2]

Let me now turn to the task of describing the alternative a charity-based theory of rationality presents. We should begin by noting that one of the most striking features of a charity-based theory is its indexical element. The similarity discerned in an attribution of rationality, according to the principle of charity, is a similarity to what *we* would find to be rational. It is not simply a matter of a similarity to some antecedently identified standard. If it were, then it would be difficult to see how the principle of charity could be anything but a convenient method for identifying rationality rather than an essential feature of rationality itself. Its use would be much like that of touch in determining temperature—a rough procedure eventually to be replaced by the development of a "rationality thermometer". This is one source of the incompatibility of a criteria-based theory of rationality with a charity-based theory, for the necessity of this indexical element rules out the possibility of there being a complete set of explicit criteria identifying what the rationality of a judgment consists in. It thus signifies a limit to how explicit a charity-based theory of rationality can be. According to such a theory, this limit reflects the claim that our standards of rationality cannot be made fully explicit. Saying that our standards cannot be made fully explicit does not reject our ordinary appeals to rules or methods for determining rational beliefs. What it does reject is that any of these rules can be given an exceptionless formulation.

---

[1] Lakatos [1978] and Laudan [1977] are good examples of philosophers who write with mainly this task of a theory of rationality in mind.

[2] The relation between the second and third tasks—those of analysis and attribution—would seem to be even closer. An adequate analysis of rationality should enable one to determine attributions of rationality. However, it is not obvious that the converse holds as well. The significance of this will emerge later.

These features immediately affect the prospects this theory of rationality has for accomplishing two of the three tasks mentioned above. It is difficult to see how a charity-based theory of rationality could accomplish the task of providing an analysis of rationality. Not only would such an analysis require identifying exceptionless rules, the need to appeal explicitly to what we would find rational would also run counter to those austere projects whose aim is that of providing an analysis in terms that do not explicitly mention rationality. Furthermore, when we turn to consider the first task of a theory of rationality—that of answering the skeptic—it becomes apparent that a theory of rationality that takes the principle of charity seriously is a complete non-starter. It cannot be employed in an answer to skepticism since its distinctive indexical feature begs the skeptic's question concerning *our* rationality. Consequently, charity-based theories of rationality are of interest if they are confined to the third task—that of settling attributions of rationality—where this task is understood to have a point independent of the considerations of skepticism underlying the first task or the considerations of reductionism underlying the second. This may not reflect an undesirable limitation on charity-based theories since each of these considerations is quite controversial.

Nevertheless, Davidson has given an argument that seems to challenge part of this claim, for he argues that the principle of charity can be used to answer the skeptic. According to Davidson's argument,

> . . . there is nothing absurd in the idea of an omniscient interpreter; he attributes beliefs to others, and interprets their speech on the basis of his own beliefs, just as the rest of us do. Since he does this as the rest of us do, he perforce finds as much agreement as is needed to make sense of his attributions and interpretations; and in this case, of course, what is agreed is by hypothesis true. But now it is plain why massive error about the world is simply unintelligible, for to suppose it intelligible

is to suppose there could be an interpreter (the omniscient one) who correctly interpreted someone else as being massively mistaken, and this we have shown to be impossible.[1]

The crucial step in this argument is the move from the claim that the omniscient interpreter's theory of our language is one in which there is much agreement between him and *us interpreted according to his theory* to the claim that his theory of our language is pretty much the same as our theory of our language. That is to say, nothing in this argument so far rules out the possibility that the omniscient interpreter's theory requires an interpretation of our utterances and actions that is quite different from our own. This, from our standpoint deviant, interpretation may be the result of his finding as much agreement as is needed to make sense of us. Consequently, even if his argument shows that there is an interpretation of us according to which we cannot be massively mistaken, it doesn't show that this interpretation may not be quite different from our interpretation of ourselves. Ruling out this possibility is precisely the skeptic's question.

The interest in a charity-based theory of rationality lies, as I have said, in its ability to accomplish the third task—that of settling attributions of rationality. However, it might be thought that there are problems even here. Doesn't this view, with its essential appeal to what we would find rational, embrace an objectionable relativism regarding rationality? In order to determine whether this is so, we need to state clearly the thesis of relativism in its most general form.

Relativism about a certain class of judgments is the view that the truth or falsity of these judgments depends upon a factor that is not explicitly mentioned in these judgments. With one exception, what makes relativism a controversial issue is not the mere claim of relativism or dependence itself. When controversy arises, it does so just in so far as the appropriateness or relevance of the alleged factor is in dispute. For example, moral relativism is generally held to be the claim that the truty or falsity of moral judgments depends upon cultural factors. What makes this claim controversial is that moral judgments are arguably universal in such a way as to preclude dependence upon cultural factors. This controversy notwithstanding, there are other factors on which moral judgments do depend. These also generate versions of "moral relativism", but these versions are not generally regarded as controversial.[2]

The striking exception is the case of relativism about truth, for that claims, in effect, that *every* judgment must be relativized. The problem with this—one that *is* independent of what truth is alleged to be relativized to—is easily seen once we realize that the relativist's point can also be made in terms of whether a given judgment is complete. Instead of saying that the truth of "Such-and-such is good" must relativized, we can say that that judgment, as it stands, is incomplete. The complete judgment results from making the implicit factor explicit. The resulting judgment becomes something like, "Given the Polynesian culture, such-and-such is good". Seen in this way, a relativism about truth, *i.e.*, a relativism about all judgments, is the claim that all judgments are incomplete. Thus, what the truth of any judgment depends upon can never be fully stated. The (complete) claim of relativism about truth itself becomes either unsayable or self-contradictory.

---

[1] Davidson [1984], p. 201.

[2] For more discussion, see my [1985].

There are therefore a number of different theses that could be characterized as a "relativism regarding rationality". What makes any one of them objectionable turns on what the particular factor is that the truth or falsity of judgments of rationality allegedly depends upon. The mere claim of relativism itself is not an issue.

This is in sharp contrast to a view of the matter such as Newton-Smith's. Instead of characterizing relativism regarding rationality as having the form,

The truth of "R is a good reason for believing that P" depends upon F,

he characterizes it as having the form,

R is a reason for holding that p is true for A while R is not a reason for holding that p is true for B.[1]

Newton-Smith stresses that the contexts *A* and *B* find themselves must be "relevantly similar" in order to avoid trivialization. But this obscures the issue since what constitutes a "relevantly similar" context is precisely the issue between the relativist and the non-relativist. Think of the moral relativist who claims that the truth of the judgment that such-and-such is good depends upon cultural factors. His claim just is that a difference in culture is a *morally* relevant difference. That is why, according to the moral relativist, there might not be any contradiction between the European's judgment that such-and-such is good and the Polynesian's judgment that such-and-such is not good. The same point holds in the case of relativism regarding rationality. However, Newton-Smith's formulation would have the relativist both assert and deny that *A* and *B* are in relevantly similar contexts. Not surprisingly, he finds that relativism lands into a contradiction, but this contradiction reflects no more than his initial characterization.[2]

The interesting question is thus not whether a charity-based theory of rationality embraces a relativism; rather, it is whether our ascriptions of rationality should depend upon what we would judge to be rational. Put like this, the answer appears to be straightforward. Of course, *our* ascriptions of rationality depend upon what *we* judge to be rational.[1] But this no more generates an objectionable form of relativism about rationality than does the fact that our determination of the truth of some sentence depends upon what else we judge to be true generate a relativism about truth. None of this, it should be noted, depends upon taking the principle of charity seriously. As far as the question of relativism goes, a charity-based theory need not be any more or less relativistic than a criteria-based theory.

---

[1] Newton-Smith [1982], p. 110. I have changed the notation slightly.

[2] Of course, I'm not denying that there are self-contradictory versions of relativism. See above. But there is much confusion here. Witness Barnes and Bloor's [1982] conflation of relativism with skeptical considerations.

(To be sure, in contrast to a criteria-based theory of rationality, a charity-based theory does not allow for a distinction between our ascriptions of rationality and what is really rational, comparable to that between our determinations of truth and what is really true. It can, of course, recognize the mistakes, errors, oversights, confusions, and other failures of rationality on the part of individuals. It is the *global* mistakes, etc., the sort that skeptical considerations feed on, that are not accommodated. In short, a charity-based theory can recognize the failures of rationality that we can recognize; it cannot recognize the failures that we cannot recognize.)

It might be objected that even if a charity-based theory of rationality does not embrace an objectionable relativism, it does embrace a view no less problematic, a view we might call "cognitive imperialism". From the standpoint of the first task of a theory of rationality, cognitive imperialism is just another name for dogmatism, and so it renders this theory of rationality useless as far as this first task is concerned. Is this cognitive imperialism equally fatal to a satisfactory realization of the third task? In order to answer this, we need to glance briefly at the controversy over what has sometimes been called "the Whig interpretation of history".

The Whig interpretation of history refers to a jumble of claims, not all of which anyone would want to defend.[2] Its central tenet is that history is to be written from the standpoint of the present. Judgments concerning what are the significant events, which represent successes and which represent failures, what is a progressive move and what is a dead-end are to be made from our present perspective. It is largely a story of progress, which, because of its perspective, glorifies the present age by portraying it as the cutting edge. Its opposite (Tory history?) is the view that the past should be studied from that past's own perspective. Talk of progress, agency or origins should be looked upon with suspicion. When taken crudely and simple-mindedly, each view leads to bad history: Whig history can yield a history that is distorted and anachronistic, full of special pleading and propaganda; its opposite can yield a history that is nothing but mere chronology and anecdote.

Nevertheless, Whig history should not be dismissed because of its simple-minded versions. More sophisticated, and more tenable, versions of Whig history can be found. In particular, the history of science often exemplifies a sophisticated, Whig approach. Historians of science are rightly interested in more than chronology and anecdote, their researches turning more to such things as identifying the origins of modern science or understanding the episodes that constituted significant scientific progress. And significance here is often determined by hindsight. More than one historian of science has acknowledged the necessity of some element of Whiggism in the history of science.[1]

---

[1] Cf. Lewis, *op. cit.*

[2] The most well-known exposition and criticism of Whig history is Butterfield [1965].

History of science informed by a charity-based theory of rationality results in a brand of Whig history. Judgments concerning the rationality of past scientists are made from the present (*i.e.*, our) point of view. For example, whether the Medicean astronomers were rational in their refusal to believe Galileo's observations based on the telescope—*i.e.*, whether their refusal is a piece of internal or external history of science—is determined by what we would find rational to believe were we in their epistemic position. Arguably, this is how the historian actually proceeds in rendering the past intelligible. The cognitive imperialism of a charity-based theory of rationality does not undermine realizing the third task of a theory of rationality. On the contrary, it is an arguably important part of it.[2]

Let us now turn to the two arguments for a criteria-based theory of rationality mentioned earlier. As we can see, neither argument is conclusive.

The first argument claimed that the objectivity of rationality required that there be criteria of rationality, for the only alternative to that of appealing to criteria, so the argument went, is that of appealing to a subjective "sense of evidence" or "feeling of plausibility". What is allegedly objectionable about appealing to such a sense is that (a) it can vary from person to person and (b) it is often deceptive.[3] This argument can easily be met because a charity-based theory of rationality presents *another* alternative, one to which these worries do not apply.

(1) The principle of charity, and with that any charity-based theory of rationality, rests on certain facts about us, the most important being that it makes sense to talk of *our* standards of rationality and of what *we* would find rational. That is to say, the principle of charity assumes as a natural fact about us that we agree for the most part on what is rational. Without that fact, there is no *us* for that principle to appeal to. Of course, agreement for the most part does not mean consensus nor does it imply anything about the judgments of any particular individual or about the judgments of the majority of individuals. We can talk about *our* standards of scientific evidence even though most of us are not scientists.

This last point needs to be emphasized, for the principle of charity employs a use of the first person plural that has not been examined much. This is a "we" that cannot be understood as a quantification over individuals nor as a royal or editorial "we". It is the "we" found in such statements as "We now know the solution to Hilbert's Tenth Problem". Its striking feature is that this "we" does not entail an "I": Although I can correctly assert that *we* now know the solution to Hilbert's Tenth Problem, it does not follow that I can correctly assert that *I* now know the solution to Hilbert's Tenth Problem. A more precise understanding of charity-based theories of rationality will require a more detailed study of this use of the third person plural.

---

[1] See, for example Hall [1969], esp. p. 216ff, or Laudan [1984], p. 65.

[2] For a defense of this last claim, see my [1986].

[3] Hempel [1965], p. 10.

(2) The claim that our sense of evidence is *often* deceptive is surely an exaggeration. But even putting that aside and allowing for the fact that we have ways of recognizing when some particular method or procedure is significantly unreliable, all of which a charity-based theory can acknowledge, an important point nevertheless remains. A charity-based theory of rationality cannot validate our standards of rationality. But, as I have argued, its inability to reply to the skeptic does not undermine a charity-based theory of rationality. It merely puts it in its place.

According to the second argument, a criteria-based theory of rationality is required because reasoning is a rule-governed activity. In reply, it could be pointed out that assuming that reasoning is a rule-governed activity begs the question concerning the tenability of charity-based theories. Nevertheless, the matter should not be left at that. After all, it cannot be denied that that assumption is quite plausible. But what is the source of its plausibility? It is certainly not based on observation. We do not always follow explicit rules when we reason, and the occasions when we do can be accommodated by a charity-based theory. Furthermore, it is a philosophical commonplace, at least since the time of Kant,[1] that applying a rule cannot, in general, be itself a rule-governed activity. Hence, not all cases of reasoning can be seen as rule-governed activity. Nevertheless, the view does have some plausibility. Its source, I would suggest, is that reasoning is normative, and a natural way of characterizing being normative is that of being rule-governed. None of this, however, presents a problem for a charity-based theory of rationality. In projecting what we would find rational, we are also projecting the appropriate normative claims of what ought to be believed, given the others' situation.

The conclusion to be drawn from all this is that a charity-based theory of rationality has much going for it. By downplaying the importance of rules or methods in science, it is not caught in the problems of trying to formulate such rules. Its natural harmony with the demands of the third task of a theory of rationality—that of settling attributions of rationality—speaks in its favor, and its inability to meet the demands of the first or second task of a theory of rationality—that of replying to the skeptic or that of providing an analysis—will not be considered a drawback by some philosophers. Nevertheless, a criteria-based theory of rationality, if one were possible, is plainly preferable. But it has not been my intention to show that one is not possible. Rather, it has been to show that were all such theories shown to be impossible, no conclusion of skepticism, subjectivism, or irrationalism in science need be drawn.[1]

---

[1] See, for example, *The Critique of Pure Reason*, B173, as well as the Preface to *The Critique of Judgment*. Modern discussions have centered around Kripke's [1982] exposition of Wittgenstein.

Center for Philosophy and Public Policy
University of Maryland

REFERENCES

ACHINSTEIN, PETER [1978]: "Concepts of Evidence", *Mind*, **87**, 22–45.

BARNES, BARRY and BLOOR, DAVID [1982]: "Relativism, Rationalism and the Sociology of Knowledge", in Hollis, Martin and Lukes, Steven (eds.), *Rationality and Relativism*. Cambridge, Mass.: MIT Press.

BUTTERFIELD, HERBERT [1965]: *The Whig Interpretation of History*. New York: Norton.

CARNAP, RUDOLF [1962]: *Logical Foundations of Probability*. Chicago: University of Chicago Press.

DAVIDSON, DONALD [1984]: *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press.

FEYERABEND, PAUL [1975]: *Against Method*. London: New Left Books.

GRANDY, RICHARD [1973]: "Reference, Meaning and Belief", *Journal of Philosophy*, **70**, 439–452.

HALL, A. RUPERT [1969]: "Can the History of Science Be History?" *The British Journal for the History of Science*, **4**, 207–220.

HARMAN, GILBERT [1982]: "Metaphysical Realism and Moral Relativism", *Journal of Philosophy*, **79**, 568–575.

HEMPEL, CARL [1965]: *Aspects of Scientific Explanation*. New York: The Free Press.

HEMPEL, CARL [1981]: "Turns in the Evolution of the Problem of Induction", *Synthese*, **46**, 389–404.

KRIPKE, SAUL [1982]: *Wittgenstein on Rules and Private Language*. Cambridge, Mass.: Harvard University Press.

LAKATOS, IMRE [1978]: *The Methodology of Scientific Research Programmes*. Cambridge: Cambridge University Press.

LAUDAN, LARRY [1984]: *Science and Values*. Berkeley: University of California Press.

LEWIS, DAVID [1983]: "Radical Interpretation", in *Philosophical Papers*, Oxford: Oxford University Press.

MCGINN, COLIN [1977]: "Charity, Interpretation, and Belief", *Journal of Philosophy*, **74**, 521–535.

NEWTON-SMITH, W. H. [1981]: *The Rationality of Science*. London: Routledge & Kegan Paul.

NEWTON-SMITH, W. H. [1982]: "Relativism and the Possibility of Interpretation", in Hollis, Martin and Lukes, Steven (eds.), *Rationality and Relativism*. Cambridge, Mass.: MIT Press.

REICHENBACH, HANS [1938]: *Experience and Prediction*. Chicago: University of Chicago Press.

THAGARD, PAUL and NISBETT, RICHARD [1983]: "Rationality and Charity", *Philosophy of Science*, **50**, 250–267.

VAN FRAASSEN, BAS [1980]: *The Scientific Image*. Oxford: Clarendon Press.

VAN FRAASSEN, BAS [1983]: "Glymour on Evidence and Explanation", in Earman, John (ed.), *Testing Scientific Theories*. Minneapolis: University of Minnesota Press.

WACHBROIT, ROBERT [1985]: "Relativism and Virtue", *The Yale Law Journal*, **94**, 1559–1565.

WACHBROIT, ROBERT [1986]: "Progress: Metaphysical and Otherwise", *Philosophy of Science*, **53**, 359–371.

---

[1] I am grateful to Christopher Cherniak, Mark Kaplan and Thomas Ricketts for conversations on these topics.

Brit. J. Phil. Sci. 38 (1987), 49–74 *Printed in Great Britain*

Induction before Hume

by J. R. MILTON

1 *Introduction*
2 *A Historical Survey*
3 *The Varieties of Inductive Scepticism*
4 *Hacking's Account*
5 *An Alternative Explanation*

1 INTRODUCTION

The history of what philosophers and scientists have thought about induction has received surprisingly little attention. The reasons for this are not particularly obscure or problematic. Many philosophers in the analytic tradition have professed a conception of the nature of their subject which makes the history of philosophy almost completely irrelevant: the occasional remarks and discussions about historical figures are as perfunctory as those which appear in scientific textbooks, and indeed have an essentially similar function. Other traditions in philosophy may take its history more seriously, but unfortunately they are often either uninterested in or else even contemptuous of the whole subject of inductive inference. Finally, one important and highly influential school within modern philosophy of science has denied not only the rationality but even the existence of inductive reasoning. If such a view is taken seriously then a history of opinions about induction becomes not merely a potentially unprofitable but also an exceedingly problematic undertaking.

If we look back at the history of thinking about induction, two figures appear to stand out from the remainder. Francis Bacon appears, as he would have wished, as the first really systematic thinker about induction; and David Hume appears as perhaps the first and certainly the greatest of all inductive sceptics, as a philosopher who bequeathed to his successors a Problem of Induction, which might be solved, or dissolved, or by-passed, but which could not legitimately or honestly be ignored.

This conception of Hume's achievement, which can be found in the writings of so many twentieth-century philosophers, may seem fairly uncontroversial. It does however carry with it a number of interesting problems. One is that it is in fact by no means obvious that Hume intended to put forward the views which have been so frequently attributed to him

Received 3 October 1985

in this century.[1] Although Hume is now quite routinely interpreted as an inductive sceptic, anyone who reads the relevant sections of the *Treatise* or the first *Enquiry* can hardly fail to notice that the *word* 'induction' is completely absent. In fact it appears only once in the *Treatise*, and once in the *Enquiry concerning the Principles of Morals*. In neither case does Hume's employment of the word conform at all closely to modern usage. In the *Enquiry* Hume is considering what he calls the general foundation of Morals: 'whether they be derived from Reason or from Sentiment; whether we attain the knowledge of them by a chain of argument and induction, or by an immediate feeling of finer internal sense' (Hume [1751], p. 170). The use of the word in the *Treatise* is even stranger. It occurs in the Appendix—a group of miscellaneous additions to Book I tacked on to Book III when it appeared in 1740. Hume has been explaining further his theory of the nature of belief, and adds, 'I conclude, by an induction which seems to me very evident, that an opinion or belief is nothing but an idea that is different from a fiction, nor in its nature, or in the order of its parts, but in the manner of its being conceived.' This use of the word 'induction' may seem strange to modern readers, but, as we shall see, it has a good historical pedigree, and there is no reason to suppose that Hume expected his readers to be either surprised or uncomprehending.

At this point someone may say that we are not or at least should not be primarily interested in this history of the *word* 'induction'. We should be concerned with the history of the *thing*. The word may have other, irrelevant, uses—it obviously does in physics, for example—but we can and should ignore these. Our concern is with what people in the past have thought about induction, whatever the terminology they have seen fit to use.

---

[1] Another interesting problem which lies outside the scope of this paper, concerns the history of the interpretation of Hume's arguments. It would appear that until relatively recently most philosophers either did not regard Hume as an inductive sceptic, or else did not suppose him to have made any points which needed a serious reply. The nineteenth century, especially the first half, was the great age of inductive theories of scientific method, but throughout all the controversies of that period Hume's name is almost completely absent. Whewell and Mill disagreed deeply about the nature of scientific method, but neither of them appears to have thought that Hume had anything to contribute to the subject. One philosopher who has noticed this, Laudan, has gone so far as to say that, 'it is one of the wilder travesties of our age that we have allowed the myth to develop that 19th century philosophers of science were as preoccupied with Hume as we are. As far as I can determine none of the classic figures of the 19th century methodology—neither Comte, Herschel, Whewell, Bernard, Mill, Jevons, nor Peirce—regarded Hume's arguments about induction as much more than the musings of an historian. (Laudan [1981], p. 240). In fact there were a few nineteenth century thinkers who interpreted Hume in the way that has now become familiar, for example John Venn (Venn [1889], pp. 127–8). Venn appears to have been somewhat isolated however. There is no sign of the modern interpretation in Green and Grose's edition of Hume's works [1874]. It is interesting that Keynes writing *c.* 1910, felt it necessary to remark that though Hume's sceptical criticisms are usually associated with causality, induction was the real object of his attack (Keynes [1921], p. 302). The older interpretation survived as late as Kemp Smith's magisterial *The Philosophy of David Hume* (1941). It would appear that the modern interpretation of Hume as an inductive sceptic arose as a by-product of work done on inductive logic.

Up to a certain point, at least, it is difficult to dissent from this. Plato uses *epagôgê* to mean an incantation (*Republic*, 364C, *Laws*, 933D), but one does not feel oneself to be taking any great risks in neglecting these passages when considering Greek theories of inductive logic. The same is clearly true of modern discussions of electrostatic or electromagnetic induction. Some degree of caution is however necessary. If we regard modern conceptions of induction as being in some way standard or natural, and direct our enquiries towards discovering past employments of the same or closely similar concepts, then we are in danger of producing a severely impoverished kind of history. One of the main purposes of intellectual history is to try to enter as fully as we can into the mentalities of people who thought in ways very different from our own. If philosophers in the past have used words such as 'induction', *inductio* and *epagôgê* in ways that seem odd or surprising to us (for example Boyle's use of *epagôgê* for *reductio ad absurdum* Boyle [1772], vol. IV, p. 468), we should take note of this fact and attempt to pursue its implications, and not filter it out of our consciousness by using a defective method of enquiry.

2 A HISTORICAL SURVEY

How, then, did philosophers think about induction before Hume—before, shall we say, the middle of the eighteenth century? The obvious starting point is with the thought of Aristotle. Locke memorably and caustically remarked that God had not been so sparing to men as to make them barely two-legged creatures, and left it to Aristotle to make them rational; and Aristotle was not the first man, or even the first philosopher, to make use of inductive arguments. (One can find a very early use of the methods of agreement and difference in the Old Testament, in Judges vi. 36–40.) He indeed gave credit for the introduction of inductive arguments into philosophy to Socrates (*Metaphysics*, 1078b28). It was Aristotle nevertheless who was the first philosopher both to use a specific technical term (*epagôgê*) for what we call induction, and to give an account of the nature of inductive reasoning.

Aristotle's theory of science has a place for both deduction and induction. Scientific knowledge is obtained by demonstration from undemonstrable first principles, and knowledge of these first principles is in turn obtained by induction. One might expect therefore that Aristotle would have discussed deduction and induction at something like equal length. In fact his remarks about induction are fairly brief and in many respects very obscure.

There are two main places in which Aristotle discusses the theory of inductive reasoning. The first, in *Prior Analytics* II.23, is not very illuminating. It is concerned purely with induction by complete enumeration, and provides a good example of Aristotle's intermittent but regrettable tendency to use Procrustean methods in forcing other kinds of inference into syllogistic form.

The most important other place in Aristotle's writings in which the nature of induction is discussed is *Posterior Analytics* II.19. This chapter is notoriously one of the most obscure in all Aristotle's writings, and its interpretation is far from straightforward. A considerable part of its obscurity derives from the fact that Aristotle appears to slide without explanation from an account of how we acquire universal concepts (100a3–b3) to an account of how we acquire knowledge of universal truths (100b3ff). Sir David Ross assumed that Aristotle was concerned both with concept formation and with induction, and passes from the one to the other because of a close analogy between the two (Ross [1949], p. 675). Jonathan Barnes on the other hand supposes that only concept formation is involved, and that Aristotle uses *epagôgê* 'in a weak sense, to refer to any cognitive progress from the less to the more general' (Barnes [1975], p. 256). This problem and others closely related to it have recently been the subject of much discussion among specialists in ancient philosophy (Barnes [1975], Hamlyn [1976], Engberg-Pedersen [1980], Upton [1981], Kahn [1981]). Like most really well established disputes in ancient philosophy, this one is unlikely ever to be finally and definitively resolved. All the less transient interpretations have at least something to be said for them, and we have no final assurance that Aristotle ever formulated a single coherent, or even approximately coherent theory. Further minute analysis of Aristotle's Greek text is unlikely to produce much further enlightenment, indispensable as such analysis certainly is. I would therefore wish to excuse myself from attempting any direct contribution to this debate (except to note a broad agreement with Kahn's approach). Instead it would seem to me useful to look first at the uses to which *epagôgê* was put by Aristotle, and then at the subsequent history of *epagôgê* and of non-deductive inferences generally. By doing this we can hope to gain insight, not so much into what was in Aristotle's mind when he was writing the *Posterior Analytics*, as into the problems and possible solutions characteristic of any broadly Aristotelian system of philosophy.

Aristotle uses the word *epagôgê* and its derivatives with what seems at least to us to be a large variety of senses. Sometimes the meaning seems to be *experience* or *observation* (*Physics*, 185a14; *De Caelo*, 276a14), or *example* (*Physics*, 229b3). More commonly some element of generalisation is involved, but the content of the generalisations is likely to appear strange to someone familiar only with the modern tradition of inductive logic stemming from Bacon. Sometimes we have the kind of argument familiar from the Socratic dialogues: 'If the skilled pilot is the best pilot and the skilled charioteer is the best charioteer, then in general the skilled man is the best in any particular sphere' (*Topics*, 105a15–17). In the majority of cases however what is established by induction has even less claim to be considered as an empirical generalisation. Among the truths which Aristotle describes as being reached by induction we have the principle that non-accidental changes occurs only between contraries, between their intermediaries and between contradictories (*Physics*, 224b30); the principle that whatever is posterior in the order of development is prior in the order of nature (*De Partibus Animalium*, 646a30); the principle that contrariety is the greatest difference (*Metaphysics*, 1055a6); and the principle that excellence is the best position, state or capacity of anything that has some employment or function (*Eudemian Ethics*, 1219a1). What we do *not* find are what we are accustomed to think of as empirical generalisations. Aristotle uses the word *epagôgê* and its derivatives over fifty times in his various writings, and the *only* example of a proposition derived by *epagôgê* which could reasonably be described as an empirical generalisation is the discussion example of all bileless animals being long-lived which appears in *Prior Analytics*, II.23. (On the background to this example, see Guthrie [1981], pp. 194–5.) It is noteworthy that in this case Aristsotle states explicitly that the induction requires a survey of *all* the particular instances.

It appears therefore that although Aristotle's formal position was that first principles of the sciences are obtained by induction, he was not an inductivist after the manner of Bacon, or Herschel, or Mill. Drawing up empirical generalisations from a wide and varied range of particular instances played little part in his scientific practice.

Aristotle's examples of inductive inferences can therefore be divided into two classes. First we have broadly common-sense arguments, usually appearing in rhetorical contexts, whose purpose is to establish some general thesis about human life and conduct. The argument about skilled pilots and charioteers in the *Topics* is an example, and there are other specimens in the *Rhetoric* (*e.g.*, 1398b5–18). These may be termed *rhetorical* inductions. Secondly there are more abstract arguments which are intended to establish some theoretical point within philosophy. These may be called *philosophical* inductions.

If we examine the rather scanty material on induction which has survived from the time of Aristotle's successors down to the end of the ancient world, a broadly similar picture emerges. Rhetorical inductions are used by Cicero, who introduced the word *inductio* as an exact equivalent for *epagôgê* (*Topica*, 42; *De Inventione*, I.51–6), and by Quintilian (*Institutio Oratoria*, V.x.73). This rhetorical tradition continues as late as Boethius, whose *De Topicis Differentiis* was a souce of Aristotelian ideas in the early Middle Ages when all knowledge of the *Posterior Analytics* and the *Topics* had been lost (*De Topicis Differentiis*, 1183D–1184D). The other, philosophical, usage can be found in Plutarch (*Moralia*, 957C) and in Plotinus, who uses the word *epagôgê* twice, once for an argument to show that there is nothing contrary to substance (*Enneads*, I.8.6.30) and once for an argument that whatever is destroyed is composite (*Enneads*, II.4.6.10).

There are other remarks about *epagôgê* elsewhere in the Platonist tradition: very briefly in Albinus (*Didaskalikos*, 158.1) and at considerably more length, in the Middle Platonist source incorporated by Diogenes Laertius in his life of Plato. Here three types of *epagôgê* are distinguished

(III.53–5). One, called *epagôgê kat'enantiosin* is a curious (and in the example given, grossly fallacious) kind of *reductio ad absurdum* (this may be the source of Boyle's curious use of the word *epagôgê*, mentioned earlier). The other, *epagôgê ek tês akolouthias*, has two varieties. One is inference from particulars to other particulars, for example from bloodstains to murder. The other is inference from particulars to universals. Here too the universal proposition given as an example is not one which we would naturally think of as an inductive generalisation, being the principle that opposites come from opposites.

The word *epagôgê* occurs in some of the fragments of Epicurus' *On Nature*, but not, apparently, with the technical Aristotelian meaning (Sedley [1973], p. 66). The Epicureans were however some of the strongest ancient advocates of the use of induction, their preferred name being inference from similarity (*metabasis kath' homoiotêta*). Our best source for these views is the treatise *On Signs*, written in the first century BC by the Epicurean Philodemus and preserved among the Herculaneum papyri.

Philodemus' treatise, so far as we can judge from its surviving parts, was a defence of inductive and analogical inferences against various objections. The source of these objections is not clearly identified in the parts of the work which we possess, but most modern scholars attribute them to the Stoics. The Stoics were indeed quite as hostile to induction as the Epicureans had been well-disposed. One possible explanation for this is that they rejected the whole idea of rational non-deductive inference. Burnyeat ascribes to the Stoics the view that the logic of our reasoning is always deductive (Burnyeat [1982], p. 236, *cf.* p. 231). 'The upshot is that Stoic logic guarantees to Stoic epistemology that the only warrant which one proposition can confer on another is the warrant of conclusive proof' (*ibid.*, p. 235). Unfortunately the nature of the surviving evidence makes interpretation difficult and more than usually precarious. It is possible that Chrysippus was as explicit as Popper, but Chrysippus' works have all been lost, and nowhere in the surviving sources is there a clear statement of the position ascribed by Burnyeat. Moreover the Stoics' opposition to induction can be explained without supposing them to have been strict deductivists. The mere fallibility of inductive inferences would, for the Stoics, have been a powerful reason for discarding them altogether. Merely fallible inferences cannot provide us with *knowledge* of anything, for according to Stoic doctrine we only know something when we have an intellectual grasp of it which cannot be weakened by further evidence or argument. Belief or opinion, which can be so weakened, is a very inferior state of mind. Indeed the Stoic ideal, the Sage, is characterised by his refusal to hold any mere opinions; like the ideal sceptic, he lives *adoxastôs*, without beliefs of any kind.

One basic Stoic objection to inductive inferences was therefore that they are inherently insecure. We cannot survey all the individual instances, and if we survey only some we risk failing to include the kinds of exceptional case which we already know both can and do exist (*On Signs*, para. 3–4, 28).

The Epicurean reply to these arguments was that inductive inferences are trustworthy provided that we take the appropriate precautions. We should make our inferences 'from what has been tested from every side, and does not exhibit a spark or trace to the contrary' (para. 45). All men who have been beheaded, for example, die at once (para. 18). Moreover we should choose characteristics which belong to all our sample without variation: 'For example, if men are found to differ from one another in all other respects, but in this respect they have been observed to have no difference, why should we not say confidently on the basis of the men we have met with and those of whom we have historical knowledge, that all men are liable to old age and disease?' (para. 35).

The dispute between the Stoics and the Epicureans had a close parallel in the field of medicine, where the Empirical school advocated the use of inductive arguments and the Dogmatic school rejected them. Galen preserves a remarkable soritical argument against reliance on generalisations based on a multiplicity of observations. If n observations are insufficient to establish reliably the truth of a generalisation, where n = 1 or some other small number, then n+1 observations must also be insufficient. If it were the case that (say) 49 observations were not enough, whereas 50 were, then it would follow that one observation, the 50th, would in itself be sufficient, which is both implausible and contradicts the initial assumptions (Galen, *On Medical Experience*, pp. 96–7 Walzer).

This argument is interesting for many reasons (Barnes [1982]), but not least for the fact that it is not an argument which found favour with modern opponents of induction. This may in part be because *On Medical Experience* has survived only in an Arabic translation, and has therefore been effectively inaccessible prior to the publication of Walzer's translation in 1944. Another reason would be that ancient philosphers took soritical arguments far more seriously than most modern philosophers have thought it necessary to do. To most modern philosophers soritical arguments appear, at their best, to be ingenious, perhaps remarkably different to analyse properly, but at bottom fundamentally *sophistical*. Much modern discussion of scepticism is characterised by what Burnyeat has called 'insulation' (Burnyeat [1984], p. 225): sceptical doubts are not allowed to influence what we do or think outside of philosophy. One of the remarkable things about modern inductive scepticism is that sceptical doubts about induction are much less well insulated from other beliefs than sceptical doubts about time, or memory, or the external world. As a result, arguments which are felt at bottom to be merely sophistical are likely to appear more out of place in discussions of induction than they do elsewhere.

The ancient sceptics appear in fact to have had rather less to say about induction than many of their modern successors. The most substantial discussion of induction surviving from either the Academic or the Pyrrhonist tradition is to be found in ch. 15 of Book II of Sextus Empiricus' *Outlines of Pyrrhonism*:

> It is also easy, I consider, to set aside the method of induction. For, when they propose to establish the universal from the particulars by means of induction, they will effect this by a review either of all or of some of the particular instances. But if they review some, the induction will be insecure, since some of the particulars omitted in the induction may contravene the universal; while if they are to review all, they will be toiling at the impossible, since the particulars are infinite and indefinite. Thus on both grounds, as I think, the consequence is that induction is invalidated.

On the surface this passage is fairly straightforward: Sextus appears to be making the familiar point that inductive generalisations derived from an incomplete survey of the relevant particulars are insecure—the word Sextus used, ***bebaios***, means safe, steadfast, guaranteed. There is nothing here to suggest that Sextus regarded such inferences as epistemically worthless. Moreover other passages suggest that Sextus had no objection to make against inductive inferences from observables to observables, as from smoke to fire or from a scar to a wound (*PH.* II.102, *Adv. Math.* VIII.154–8). Sextus' polemic was directed against what he called indicative signs: inferences from observables to unobservables, such as Epicurean atoms or Aristotelian elements.

The real problem with understanding this passage is that we can only do so by placing it in its overall intellectual context, and that context itself is highly problematic. It is clear that Sextus assented to the use of at least some inductive inferences, but it is by no means clear what this assent amounts to. Very roughly there are two lines of interpretation: one is that Sextus is concerned merely to doubt the philosophical theories put forward by his dogmatist opponents; the other is that his doubt extended also to the assumptions and inferences made in everyday life. On the former interpretation, Sextus was only mildly sceptical about inductive arguments: they were not wholly reliable and their employment had to be restricted to the everyday world, but subject to these limitations they could legitimately be used. On the latter interpretation Sextus was a sceptic about inductive inferences, not because he found them particularly obnoxious, but because he was sceptical about all inferences without exception.

Neither of these interpretations can easily be dismissed—indeed reading the rapidly growing secondary literature on this topic can tend to induce the kind of suspension of assent which it was the main aim of Sextus' philosophy to produce. The most commonly held view until quite recently was that Sextus was a moderate who was not disposed to reject inferences from observables to observables (Stough [1969], pp. 128–39). More recently opinion has shifted in favour of Sextus as a more thoroughgoing sceptic (Burnyeat [1984], Barnes [1982a]), even though it is generally admitted that the claims that Sextus makes in various places are by no means obviously consistent with one another (for an excellent analysis of the problem, see

Barnes [1983], pp. 154–60; Stough [1984] now advocates a position close to that of Burnyeat, but she still holds (p. 155n) that Sextus is committed to a principle of induction). It does nevertheless seem reasonably clear that Sextus was not particularly hostile towards inductive arguments: on one interpretation they survive; on the other they are destroyed, but only because all arguments are destroyed.

The medieval schoolmen, who had so much to say about deductive inference, contributed relatively little to the theory of induction. Most of the logic textbooks found some space for a discussion of induction, but the remarks made were usually brief and rather perfunctory. Even Ockham's *Summa Logicae*, a treatise planned on a much larger scale than usual, and one of the most remarkable achievements of medieval logic, conforms to the same pattern. Ockham devoted six chapters of this work (III. iii.31–6) to induction, but the treatment is unilluminating; the extra space merely gave Ockham the opportunity (fatal to so many medieval authors) of being prolix and rather pedantic. Aquinas also had little of any value to say: his remarks in his Commentary on the *Posterior Analytics* are very dull. A rather more interesting discussion can be found in the longer of Duns Scotus' two commentaries on the *Sentences* of Peter Lombard, the work now known as the *Opus Oxoniense*. Scotus raised the question which was to be central to so many subsequent discussions: can we be certain about any universal conclusions reached by means of induction? His answer was that we can, provided that we make use of general propositions not dependent on induction—in this case the principle that whatever happens in many cases as a result of a cause which is not free is the natural effect of that cause (*Opus Oxoniense*, I, d.iii, q.4, Wolter [1962], pp. 109–10). Like most if not all such general principles, this one seems both highly dubious and far from obviously capable of doing the job intended for it, but it no doubt seemed satisfactory enough to Scotus, and he was certainly to have many successors in the centuries following.

The most substantial and most influential discussion of induction in the seventeenth century is to be found in Francis Bacon's *Novum Organum*. Bacon was the first philosopher to consider induction as the chief method of inference in the natural sciences, and subsequent estimates of his philosophical stature have reflected very accurately the high or low esteem in which inductive methods have been held. The main elements of Bacon's views are well known, at least in broad outline, and need not be described here, but there are two points which perhaps need to be given some degree of emphasis.

The first is that Bacon had an extremely low opinion of induction by simple enumeration. The language he uses makes his scorn absolutely clear. Induction by simple enumeration is 'utterly vicious and incompetent' (*De Dignitate et Augmentis Scientiarum*, V.2), 'gross and stupid' (*ibid.*) and 'childish' (*Novum Organum*, I.105). Bacon's fundamental objection to this kind of induction is that it can lead us only to conjectures, and not to certain knowledge. 'For the induction of which the logicians speak, which proceeds by simple enumeration, is a puerile thing; concludes at hazard; is always liable to be upset by a contradictory instance; takes into account only what is known and ordinary; and leads to no result' (*Magna Instauratio*, Distributio Operis, Bacon [1859], vol. IV, p. 25). Modern introductory books on the philosophy of science sometime commence with a criticism of 'naive inductivism' and carefully warn us of the dangers of concluding that all swans are white from the basis of a limited number of observations. Bacon regarded all such inductions as quite useless.

Secondly, Bacon's own method of exclusion presupposed a considerable body of metaphysics. One instance of this is that Bacon assumed a simple one-to-one relation between the observable natures of bodies and the forms which are their causes. Another is that he assumed a kind of Principle of Limited Variety: that the number of different forms to be found in nature is manageably finite. Both of these assumptions are necessary if Bacon's method of induction is not to be exposed to exactly the same kind of criticisms which he himself had made of the old and despised method of induction by simple enumeration.

Gassendi's remarks about induction are not always easy to interpret. In his early sceptical phase, represented by the *Exercitationes Paradoxicae Adversus Aristoteleos*, he appears to be quite dismissive of the Aristotelian theory of induction. His reasons for this are very similar to those put forward earlier by Sextus; we cannot enumerate all the particular cases since they are actually or potentially infinite in number. Gassendi's one novelty, as compared with Sextus, was to deploy a characteristically voluntarist argument based on God's infinite power. Even if there exist some individuals which are unique, such as the sun, there are nevertheless also infinitely many possible suns which God could create, and any genuinely well-grounded propositions would have to be true of these also (*Exercitationes*, II.v.5, Gassendi [1972], p. 75).

In Gassendi's *magnum opus*, the *Syntagma Philosophicum*, his views on induction appear rather ambivalent. His main discussion, taken by itself, might suggest that the only inductive arguments of any value are enthymemes—arguments which can be turned into regular syllogisms by the addition of a concealed premise or premises. Without the addition of such a premise the inference has no force, *consequutionis vis nulla foret* (*Syntagma*, I.iii.11, Gassendi [1658], vol. I, p. 113). If all the singulars are not included in the enumeration, the proof dissolves (*labefactet*). This is a requirement which Gassendi quite freely admits to be 'very difficult, or impossible' to fulfil.

This passage might suggest that Gassendi saw inductive arguments as being of very restricted value. Elsewhere in the same work however he appears to give induction a substantially more positive role. In Gassendi's opinion the usual descriptions of arguments from the more general to the particular as *a priori*, and from the particular to the general as *a posteriori*, ought really to be reversed. It is particulars which are known first, and all the evidence and certainty which a general proposition can have is dependent on an induction from particulars (*Syntagma*, I.iii.16). This is true even of such highly general propositions as 'every whole is greater than any of its parts'. Gassendi regarded this as an inductive generalisation, based on observations made from childhood onwards that a whole man is bigger than his head, a whole house larger than a single room, and so on.

Other seventeenth-century writers were less ambiguous in their attitude toward inductive inferences. The authors of the Port Royal *Logic* for example regarded all inductions based on a survey of fewer than all the relevant particular as merely sophistical:

> Induction is not at all a certain means of knowing something, except when we are sure that the induction is complete; there being nothing more common than to discover the falsity of what we had believed to be true on the basis of inductions which had seemed so general that one would never imagine that one could find an exception.
>
> Thus until two or three years ago it was believed to be quite indubitable that when water was contained in a vessel with curved sides, one end being narrower than the other, it remained completely level, being no higher in the smaller end than in the larger one. One was assured of this by an infinity of observations.
>
> Nevertheless it has been found recently that this is quite false when one of the ends is extremely narrow, for in such cases the water rises higher in this end than in the other. All this shows us that inductions alone cannot give us a full certainty of any truth—unless we were sure that the inductions were complete, which is impossible.[1]
>
> (Arnauld and Nicole [1662], pp. 316–17)

This inherent fallibility of all inductive arguments was not a source of great worry to Arnauld and Nicole, who believed that scientific knowledge was to be obtained by deduction from self-evident and indubitable axioms. Leibniz's attitude towards induction was rather more complex. In some places he expressed views very close to those of the Port Royal *Logic*. In a letter of 1702 to Queen Sophia Charlotte of Prussia he maintained that induction can never teach us truths which are fully (*tout à fait*) universal (Leibniz [1969], p. 551). Even if we have seen one hundred times that lumps of iron sink when placed in water, we still cannot be sure that this must always happen (*ibid.*). The same is true of other inductive generalisations, such as that heavy bodies will fall, or that we ourselves will eventually die. In the former case, Leibniz commented that we cannot go with complete confidence (*bien seurement*) beyond the experiences we have had, unless we are aided by reason (*ibid.*, p. 550). In the latter, the similarity which exists between men would not by itself justify us in considering the conclusion as certain (*ibid.*, p. 551).

This view that unsupported inductive inferences cannot provide grounds for certainty appears much earlier in Leibniz's writings. In his Preface to

---

[1] The observations of capillary attraction which Arnauld refers to are probably those of M. Thévenot, undertaken in 1658–61. Cf. *Dictionary of Scientific Biography*, vol. XIII, p. 336.

Nizolius' *De Veris Principiis* [1670] he put forward the view that induction cannot produce any kind of certainty, even moral certainty, unless one also makes use of universal propositions derived purely by reason (*ibid.*, pp. 129–30). With the aid of these universal propositions moral certainty can be attained, even though perfect certainty cannot. Mere inductive generalisations cannot be certain at all.

There are places in the *Nouveaux Essais* where Leibniz appears to have been expressing the same attitude. For example, he remarked that 'however many instances confirm a general truth, they do not suffice to establish its universal necessity' (Leibniz [1765], p. 49), and he illustrates this with one of Popper's favourite examples: that the sun does not rise every 24 hours in the polar regions. There are however other passages which appear to say something rather different. In IV.vi.8, where Leibniz is arguing against Locke's pessimistic analysis of our capacity to know universal truths about substances, he remarks that

> We know almost as certainly that the heaviest of all bodies known on earth is fixed i.e., not decomposed by heating, as that the sun will rise tomorrow. This is because it has been experienced a hundred thousand times. It is a certainty of experience and fact, even though we do not know how fixity is linked with the other qualities that this body has (pp. 404–5).

Here Leibniz's view is that observation of regularities can produce certainty. 'For it seems to me that, in the case of propositions which we have learned from experience alone and not by the analysis and connection of ideas, we rightly attain to *certainty* (moral or physical, that is) but not to necessity' (*ibid.*, p. 406). Apparently similar views can be found elsewhere. In a fragment dated by Couturat to 1693 Leibniz distinguishes three levels of epistemic security: *certitudo logica*, *certitudo physica*, and *probabilitas physica*, and he makes it clear that induction, when properly carried out, produces the second of these (Leibniz [1903], p. 232). This also appears to be the implication of a remark in the *Nouveaux Essais* in which Leibniz explains physical necessity as 'necessity founded on induction from what takes place in nature' (Leibniz [1765], p. 499). It is by no means obvious that all these passages can be reconciled with one another. On the other hand the difference between the various views expressed is perhaps not all that great. Inductive arguments cannot produce the kind of certainty characteristic of mathematical demonstrations, but they can make their conclusions probable, even highly probable. Whether these very high probabilities amount to moral certainty is something about which Leibniz appears to have had no settled opinion.

Apart from Bacon, none of Hume's British predecessors has very much to say about induction. Hobbes had little time for the kind of experimental philosophy which Bacon had advocated and which some of his younger contemporaries were trying to pursue. He mentioned induction in only a few places: twice in *The Whole Art of Rhetoric* (ii.21,24), and once in the

Examinatio et Emendatio Mathematicae Hodiernae (Hobbes [1839], vol. IV, p. 179). The *Rhetoric* is a work conceived on very traditional lines, and induction appears to be included primarily because it is one of the traditional types of rhetorical argument. The remark in the *Examinatio* is slightly more informative. Hobbes's target here was John Wallis's *De Arithmetica Infinitorum*, and Hobbes objected strongly to Wallis's use of induction in arguing towards theorems about infinite series and infinite continued fractions. Induction, says Hobbes, is not demonstration '*nisi ubi particularia omnia enumerantur*'. Wallis, it may be noted, himself had exactly the same opinion of the limitations of incomplete induction: it can arrive at a conclusion which '*conjecturalis tantum est, aut probabilis, non omnino certa*' (Wallis [1687], p. 170).

Locke had slightly more to say. His attitude towards induction emerges clearly in what is to the best of my knowledge the only place in all his writings in which the word appears, section xiii of *The Conduct of the Understanding*. Locke remarked that 'those seem to do best, who taking material and useful hints, sometimes from single matters of fact, carry them in their minds, to be judged of, by what they shall find in history, to confirm or reverse these imperfect observations: which may be established into rules fit to be relied on, when they are justified by a sufficient and wary induction of particulars' (Locke [1823], vol. III, p. 214). Locke was not in any way opposed to the use of non-demonstrative arguments, whether inductive or analogical, but he insisted that the conclusions of such arguments cannot be knowledge. 'Possibly inquisitive and observing men may, by strength of *judgement*, penetrate farther, and on probabilities taken from wary observation, and hints well laid together, often guess right what experience has not yet discovered to them. But this is guessing still; it amounts only to opinion, and has not that certainty which is requisite to knowledge' (*Essay*, IV.vi.13).

This sharp distinction between knowledge and opinion or belief, and the insistence that knowledge, in order to be knowledge, must be certain, can be found repeated in many of Locke's other works. When he was forced to defend his position, it was not against anyone who supposed that knowledge could be uncertain, but against theologians like Stillingfleet and Jonas Proast who held that we could be certain about matters of religious faith (Locke [1823], vol. IV, pp. 143–7, 271–99; vol. VI, p. 558).

3 THE VARIETIES OF INDUCTIVE SCEPTICISM

From this rather brief survey one thing at least is clear. In the centuries before Hume, and especially during the 120 years which separated the *Novum Organum* and the *Treatise of Human Nature*, very few philosophers had been entirely unaffected by doubts about the reliability of inductive inferences. Moreover the doubts felt were of very different kinds: a mere division into sceptics and anti-sceptics would be far too crude to be of any real use. What we need is a more elaborate classification of the different levels of unease about inductive reasoning which have felt:

1. There are reservations about inductive reasoning which arise merely because inductive arguments are not deductively valid.
2. There is the view that inductive arguments are inherently and irredeemably fallible: although such arguments may make their conclusions probable, they can never make them certain .
3. There is the view that genuinely universal propositions can never be given a probability greater than zero by any inductive argument.
4. Finally, there is the view that no inductive arguments, whether to particular or to general conclusions, can be given any rational foundation whatever.

These views, which constitute successive and increasingly radical stages of doubt about induction, need to be kept clearly separate if confusion is to be avoided.[1]

Despite a few rather disingenuous claims to the contrary (*e.g.*, Hume [1745], pp. 19, 22), Hume clearly holds position 4, which may be termed radical inductive scepticism (Hume [1739], pp. 267–8; [1748], p. 41).

The third position, that all universal generalisations have a probability of zero given any finite quantity of evidence, has been the subject of much discussion among modern philosophers concerned with inductive logic and probability theory. It would seem however to be a view which would only be likely to appear after the development of a mathematical calculus of probabilities. I am not aware of anyone who held it in the period before Hume.

In the seventeenth century and earlier, the most commonly held position was the second—that inductive arguments are inherently fallible and produce (at best) only probability and not certainty. This was a more radical conclusion than it might seem at first sight to someone familiar primarily with twentieth-century discussions of this topic. For nearly all philosophers in the period before Hume knowledge meant certain knowledge. This was a view shared by epistemological optimists like Bacon and Descartes, and by sceptics like Foucher and Bayle. Hume himself held that 'knowledge and probability are of such contrary and disagreeing natures, that they cannot well run insensibly into one another' (Hume [1739], p. 181). That the same conception of the nature of knowledge should be shared both by thinkers as confident in the power of the human mind to acquire knowledge as Bacon and Descartes were and by sceptics like Bayle and Hume is in reality not at all strange. In the Ancient world one can find many of the same premises shared by the Stoics and their Academic and Pyrrhonist opponents. Both parties had an identical though opposite interest in insisting that what can count as knowledge must satisfy the most stringent criteria. What is more remarkable is that the same view can be found in more cautious and less confident thinkers who nevertheless had no wish to be included among the sceptics. Locke's views have already been described. John Wilkins in his classification of the kinds of assent which we can give to propositions, placed them all under two main headings 'knowledge or certainty' and 'opinion or probability' (Wilkins [1675], p. 5). Similar views can be found in many of the other philosophers and scientists connected with the early Royal Society (Hooke [1705], p. 330; Glanvill [1676], p. 45).

---

[1] Popper does not always appear to do this. For example, when he describes Hume as having produced 'a gem of priceless value for the theory of objective knowledge: a simple, straightforward, logical refutation of any claim that induction could be a valid argument, or a justifiable way of reasoning' (Popper [1972], p. 86) there appears to be a slide from position 1 to position 4. In other places Popper sometimes associates Hume with position 1 (Popper [1959], p. 312) and sometimes with position 4 (Popper [1963], p. 200). The adoption of radical inductive scepticism (position 4) has exceedingly disturbing consequences, which Hume saw more clearly than Popper has. If all arguments other than purely deductive ones have no rational foundation, then Hume's and Popper's arguments against induction must be purely deductive. Nevertheless it is quite apparent that the kind of controversy which they have provoked is quite unlike anything which indisputably deductive arguments produce, even when the conclusions of those arguments are profoundly surprising or subversive. The proofs of Gödel's incompleteness theorems are much longer and apparently more complex than Hume's arguments against induction, but they never became a subject of controversy. If both arguments are purely deductive, why should this be?

The idea that there could be knowledge which was uncertain or merely probable is much more difficult to locate. Gassendi once remarked in passing that just as one can speak of certain knowledge and certain opinion, so one can speak of fallible (*imbecillam*) knowledge and fallible opinion (Gassendi [1658], vol. III, p. 206b). The suggestion was not developed, however, and it would be imprudent to try to draw from it views which Gassendi would probably have repudiated. Leibniz held that we can have knowledge of probabilities, but by this he meant that we can know the truth of probability statements (Leibniz [1765], p. 373). There is no suggestion that we can have knowledge when we are uncertain of the truth of the propositions themselves.

The main way in which the requirement that knowledge must be certain was made less constricting was by the introduction of a variety of different types of certainty. The distinction between absolute, mathematical or metaphysical, certainty and moral certainty appeared early in the seventeenth century[1] and rapidly came into general use. In England a variety of sometimes quite complex systems of classification were drawn up. Glanvill and Wilkins both made a distinction between infallible and indubitable certainty, the former being the higher grade (Glanvill [1676], pp. 47–50; Wilkins [1675], pp. 8–10), and Stillingfleet found it useful to distinguish no less than five different degrees of certainty: metaphysical, rational, physical, infallible and moral (Stillingfleet [1710], vol. VI, pp. 86–9). Perhaps the clearest system of classification was that used by Boyle: there are three levels of certainty and three types of demonstration: metaphysical demonstrations, which presuppose nothing and give absolute certainty; physical demonstrations, which assume as true various physical principles; and moral demonstrations (Boyle [1772], vol. IV, pp. 42, 182). No-one supposed that the conclusions of inductive inferences could claim any of the more stringent grades of certainty, but if they could aspire even to moral certainty they would thereby become possible objects of knowledge. Unfortunately this is an issue about which much less is said than one might desire, mainly because most of the discussions of the grades of certainty occur in theological contexts in which the reliability of inductive inferences is not really relevant to the points at issue. Among those philosophers who did give serious thought to the use of inductive generalisations in natural philosophy there was no real agreement. The general consensus of the scholastic philosophers had been that no merely probable propositions, however great their probability might be, could be certain (Smiglecki [1638], p. 661). This view was maintained by Locke, by John Wallis (Wallis [1678], pp. 170–1), and by Glanvill (Glanvill [1676], p. 45). Other thinkers held that we could be morally certain about at least some conclusions of inductive inferences. Samuel Parker believed that all general axioms (*e.g.*, that the whole is greater than its parts) 'are only the results and abridgements of a multitude of single Experiments' and are yet 'obvious and apparent Certainties' (Parker [1666], pp. 55–6). Wilkins, more cautiously, held that we could be morally certain that the sun would continue to rise (Wilkins [1675], p. 10). Leibniz, perhaps unconsciously, seems to have vacillated between the two positions. Isaac Barrow appears equally indecisive: on the one hand the confirmation of any proposition by frequent experiments is 'almost sufficient' to enable us to consider it as universally true (Barrow [1734], p. 74), which suggests that such conclusions are merely highly probable; on the other hand when any proposition is found agreeable to constant experience 'it will at least be most safe and prudent to yield a ready assent to it' (*ibid.*, p. 73), and this would appear to imply that such conclusions are at least morally certain.

---

[1] Its origins are surprisingly obscure. Henry van Leeuwen traces it back to Chillingworth and hence (rather tentatively) to Grotius (Van Leeuwen [1970], pp. 21–2). Barbara Shapiro on the other hand ascribes it to the scholastics, though without giving any references to any scholastic author (Shapiro [1983], p. 84). Shapiro's derivation seems slightly more probable: Descartes, who was quite familiar with the idea, seems to have attributed it to the scholastics (letter to Mersenne, 21 April 1641, Descartes [1970], p. 99).

4 HACKING'S ACCOUNT

We can therefore ask ourselves two questions about inductive scepticism and the origins of the 'problem of induction':

(1) Why does *radical* inductive scepticism seem not to have appeared before Hume?

(2) Why were many earlier philosophers relatively unworried by the implications of inductive fallibilism?

One answer to the first of these questions has been supplied by Ian Hacking in his book *The Emergence of Probability*. Hacking's book is mostly, as its title suggests, about the evolution of various notions of probability, but the final chapter is explicitly about the emergence of the modern problem of induction.

Hacking makes a sharp distinction between what he calls the *analytic* problem of induction—distinguishing good and inductive reasons and classifying the various degrees of evidential support—and the *sceptical* problem (Hacking [1975], p. 176). Discussion of the analytic problem goes back at least to Leibniz and Jakob Bernoulli, but radical sceptical doubts about induction appear only with Hume (*ibid.*, p. 177). Hacking's thesis is that the sceptical problem of induction became a possible problem only as a result of two events. The first was the emergence of a concept of what Hacking calls 'internal evidence'—that is, evidence other than testimony. It was this that enabled the modern concept of probability to emerge, and with it the analytic problem of induction. The appearance of the sceptical problem required one further change. 'Once the concept of internal evidence was established by 1660, the final transformation needed for the sceptical problem of induction was this transference of causality from knowledge to opinion' (*ibid.*, p. 180).

Hacking's claim that there was no concept of internal evidence in Medieval or Renaissance Europe has been damagingly criticised by a number of writers (Blackburn [1976]; Garber and Zabell [1979]; Laudan [1981], pp. 72–85). One particular weakness is that Hacking bases his argument on a very implausible (indeed quite unsustainable) claim about natural and conventional signs. Hacking writes that:

> Arbitrary and conventional signs are carefully distinguished in the Port Royal *Logic*, the same book from which I took my terminology of internal and external evidence. Hobbes also very sharply distinguishes 'arbitrary' and 'natural' signs. Once natural signs have been distinguished from any sign of language, the concept of internal evidence is also distinguished.
>
> (Hacking [1975], pp. 47–8)

According to Hacking, therefore, the distinction between arbitrary and natural signs and the concept of internal evidence emerge together around the middle of the seventeenth century. One fundamental objection to this argument is that the distinction between natural and conventional signs, far from being a new discovery of the seventeenth century, was a commonplace of medieval philosophy. One of the standard features of scholastic treatises on logic is a section on signs and on the difference between natural and conventional signs. Indeed the distinction between these two kinds of signs is very much older still. Book II of St Augustine's *De Doctrina Christiana* is concerned with signs in general, and Augustine prefaces his discussion with a careful explanation (chs. 1–2) of the difference between those signs which signify by nature and those which signify by convention.

A rather less explicit but quite recognisable statement of the same distinction can be found in Aristotle, in chapter 2 of *De Interpretatione*, and Aristotle was almost certainly writing with Plato's *Cratylus* in mind. One of the speakers in that work, Hermogenes, maintains quite clearly the thesis that all names signify solely by convention (383C–D), and it would appear that such views are widely held among the sophists. Hacking makes much of a contrast between Paracelsus and his theory of natural signatures and Gassendi, and adds that with the work of Gassendi and his like-minded contemporaries, 'the discovery that all names are conventional thunders into modern philosophy' (Hacking [1975], p. 4). The difference between Paracelsus and Gassendi is certainly real enough, but the idea that we have here a radical *historical* discontinuity between two quite alien modes of thinking is quite illusory. The idea that a knowledge of the real name of a thing gives one an insight into its essence or nature is of incalculable antiquity, and the idea that Adam was able to give things their natural names, and not merely conventional ones, appears as early as Philo (*De Opifico Mundi*, 148–50). Such views continued to be maintained by Jakob Boehme, John Webster and others well into the seventeenth century (Aarsleff [1982], pp. 60–61). Indeed what characterises that century is not the appearance of the 'modern' view that all linguistic signs are conventional, but the effective disappearance of the opposite view as a serious intellectual option.

Hacking's other precondition for the emergence of a sceptical problem of induction, that causation must cease to be a possible subject of belief or opinion only, is more difficult to evaluate. One problem is that here, as elsewhere in the book, it is not wholly clear precisely what Hacking is trying to say. To say that 'Hume can begin only when causation is stolen from knowledge' (Hacking [1975], p. 181), or that (for Leibniz) 'Truth is ultimately demonstration' (*ibid.*, p. 185) is to indulge in a kind of philosophical impressionism: the general drift may be clear enough, but the particular point being made is not. Hacking's view appears to be that the sceptical problem of induction became possible once the old hope that one could (in principle at least) demonstrate the existence of causal connections between things had been abandoned. Hacking sees this hope fading rapidly in the late seventeenth and early eighteenth century, and finally vanishing with Berkeley, who might well have anticipated Hume had he not possessed such a strong aversion for all forms of philosophical scepticism. For Berkeley and Hume there are no necessary connections in nature. For Berkeley there are no causal connections whatever in the natural world; for Hume there are, but the idea of necessary connection comes from within ourselves, not from outside. The implication of this is that the causes of the appearance of radical inductive scepticism are ontological.

That this is Hacking's view appears most clearly in a comment made in a subsequent paper on this same topic. 'There is a sceptical problem of induction, not because (as with Glanvill) we may be in doubt that we have located the necessary connections that will guide our predictions about the future, but because we now think that there are no necessary connections, not even unknown ones' (Hacking [1981], p. 116).

This is in many ways an attractive suggestion. Glanvill's mild scepticism about his contemporaries' claims to have any kind of scientific knowledge does contrast strongly with Hume's vivid portrayal of his own complete cognitive disorientation in the last few pages of Book I of the *Treatise*. 'Where am I, or what? From what causes do I derive my existence, and to what condition shall I return? Whose favour shall I court, and whose anger must I dread? What beings surround me? and on whom have I any influence, or who have influence on me?' (Hume [1739], p. 269). There is nothing like this in any of Glanvill's writings. Glanvill indeed was much closer to Locke: confident of living in an ordered and intelligible world, but pessimistic about our capacity to discover very much of that order. Hume by contrast lacked this kind of ultimate assurance. The only feelings of confidence which he could have were those which arose from surrendering to natural propensities to believe which were themselves incapable of any kind of rational justification.

There are however some problems with this account. It could for example be objected that Hume does not deny the existence of necessary connections. If he had wanted to do so he would have denied that we have any such idea; in fact, his concern was to discover the impression from which that idea was derived. In reply to this it could be said that what Hume denied was the existence of necessary connections *in nature*: the impression from which the idea is derived is the customary transition which occurs in our thought from the idea of the cause to the idea of the effect.

That Hume held that there were no necessary connections in nature is uncontroversial. What is less clear is whether Glanvill (or indeed many of Hume's other predecessors) thought that there were. If they did not, then the radical difference between Glanvill and Hume which Hacking rightly remarks on will need to be explained on other grounds.

In considering this issue it is essential to distinguish carefully between *necessary* connections and *causal* connections. The view that there are no causal connections in nature appeared long before Berkeley or Hume. It can be found in certain Islamic philosophers such as al-Ashari (d. 935), whose views became known in the West through attempted refutations by Averroes and Maimonides (Wolfson [1976], ch. 8). Similar ideas reappeared in the seventeenth century, most notably in Malebranche. Hume was aware of at least the later stages in the history of this theory, and regarded it with little favour (Hume [1748], p. 73n). He was himself quite ready to ascribe causal powers to bodies. Indeed he had no real alternative: all the philosophers who had denied the existence of causes in nature had made God the immediate cause of all phenomena, and this was hardly an approach likely to appeal to Hume.

It would appear therefore that a denial of the existence of causal connections in nature has no obvious tendency to generate inductive scepticism. Malebranche and Berkeley were very little interested in the problems that were to concern Hume so much, and there is no sign that either of them would have had any leaning towards or any sympathy for Hume's approach.

In the case of the existence of necessary connections the situation is more complex. After all, what precisely *are* the necessary connections which Hume denied but which presumably some at least of his predecessors had supposed to exist? The fundamental maxim on which Hume grounded his argument against the existence of necessary connections is that every real thing can be supposed, without absurdity, to be capable of existing separately from every other real thing (Hume [1739], pp. 79–80, 173, 233, 247, 249, 466). It is this that prevents us from deducing any causal relations *a priori*, and compels us to reply solely on experience.

On this account there is a necessary connection between two entities a and b if either a cannot exist in the absence of b, or b cannot exist in the absence of a. Hume denied the existence of such necessary connections, but so also did a good number of his predecessors. Indeed the view that there are no necessary connections in nature resembles the view that there are no causal connections in having a past which extends back to the Middle Ages. Ockham introduced into philosophy the notion of what he called a *res absoluta*, and maintained that every *res absoluta* can exist independently of every other (*Quodlibetae*, VI, q.1; Ockham [1980], p. 605). For Ockham all substances and all real qualities are *res absolutae*.

The idea that lay at the heart of Ockham's views—that God, being omnipotent, could bring into being any possible state of affairs—was very widely held in the seventeenth century. Clearly it was an idea with the most far-reaching consequences, potentially at least. That its actual consequences were less radical can be explained partly by the general human tendency not to follow lines of argument as far as they can go, and partly for two other reasons.

The first is that almost all seventeenth-century philosophers either were positively attached to the metaphysics of substance and attribute, or at least (like Locke) found it impossible to discard it completely. Such a metaphysics clearly had a tendency to limit the scope of voluntarist arguments from divine omnipotence (as one may see if one tries to imagine how Descartes would have answered the question of whether God could create a thought without a thinking substance).

Secondly, nearly all seventeenth-century thought about the natural world had in the background the idea that the world is divinely governed. It might be admitted that God had the absolute power (to use the convenient scholastic term) to do anything that might be described without contradiction, but most of the range of possibilities thus disclosed were of little relevance to natural philosophers who believed themselves to be investigating the workings of a world created and providentially governed by a rational and benevolent deity.

Neither of these beliefs was shared by Hume, who discarded completely the traditional metaphysics of substances and attributes, and who disbelieved in the existence of God. It might be supposed that Hume's atheism would have led him to reject the voluntarist arguments against necessary connections which had relied for their force on the doctrine of divine omnipotence. In fact the conclusions survived, and the arguments were adapted with God being replaced by the human imagination. 'Whatever is clearly conceiv'd may exist after the same manner' (Hume [1739], p. 233). Hume was therefore able to reason along the same lines as his late medieval and early modern predecessors, and, unencumbered by many of their presuppositions, to arrive at conclusions which even he found to be deeply disturbing.

5 AN ALTERNATIVE EXPLANATION

Hacking's explanation of the emergence of inductive scepticism would seem therefore to be fundamentally unsatisfactory. Of the two parts of his explanation, one (the claim about signs) is certainly false, and the other is seriously inadequate. It would appear that what distinguished Hume from his predecessors was not the adoption of any radically novel metaphysical axioms. The most fundamental premises of his philosophy were far from new. What distinguished Hume and enabled him to formulate a kind of scepticism without clear historical precedent was a greater readiness and ability to pursue certain lines of argument to their ultimate conclusion, a temperament sympathetic to the construction of a systematic kind of philosophy (unlike Bayle), and a notable freedom from many of the philosophical and theological constraints which guided most of his predecessors.

In the remainder of this article I would like to set out an alternative hypothesis to Hacking's. Superficially it might appear to diverge from the account of Hume's position given in the last few pages, but on deeper examination the two analyses will be found to be entirely compatible with one another.

The problem of induction is at bottom a problem about inference from particular to universal propositions. It would seem therefore reasonable to suppose that there may be some usefully close connection between this problem and the metaphysical problems about the nature and existence of universals which have become known as the problem of universals.

The problem of universals as it existed in the Middle Ages and subsequently was essentially the problem of whether there are any *things* which are universals, or whether the only things that exist are particulars. Towards the end of the Middle Ages the two contending parties became known respectively as realists and nominalists, a pair of rather inappropriate names which have survived, but which are potentially very misleading. The view which has become known as nominalism is, despite the unfortunate connotations of the word itself, not primarily about names at all. Nominalism is best understood as the thesis that everything which exists is an individual or a particular, and realism as the denial of this. When Berkeley, in the person of Philonous, remarked to Hylas that 'it is an universally received maxim, *that everything which exists, is particular*' (Berkeley [1713], p. 192) he was both indicating his own adherence to the nominalist side and at the same time presuming that those whom he was trying to persuade would be of the same point of view. If we assume that Hylas was intended to be a plain man who had derived his philosophical opinions largely from Locke's *Essay*, then Berkeley was clearly justified in his approach. Hobbes, Locke, Berkeley and Hume were all quite unambiguously nominalists, not in the sense that they used this name for themselves, but in that they were in full agreement with Ockham and his followers on the fundamental principle that only individuals exist. (On the importance of nominalism in Locke's thought, and the misconceptions which arise when the character and historical continuity of the nominalist tradition is inadequately understood, see Milton [1981].)

The connection between the problem of universals and the problem of induction appears if we consider the truth-conditions of universal propositions. If we assume a correspondence theory of truth, then if we suppose that universal *things* of some kind exist, universal propositions can be true or false according to whether they correctly state how things are with the corresponding universal entities. This is perhaps clearest if we consider the early, as yet uncomplicated, theory of Ideas put forward by Plato in dialogues like the *Phaedo* and the *Republic*. Universal propositions about lines and circles are true not because of their relation to particular, sensible lines and circles, but because of their relation to the Ideas. Parallel accounts can be given for Neoplatonic, Aristotelian and Neo-Aristotelian theories, though the relations are more complex and more difficult to describe.

If on the other hand we accept the nominalist position, all this talk of universal entities is wholly and utterly mistaken, and so therefore are all theories which presuppose their existence. A universal proposition has to be considered as an infinite number of singular propositions. The universal proposition is true only if all the singular particular propositions are true, and it can be *known* to be true only either if it is deduced from higher-level universal propositions which are themselves known to be true, or else if we know that *all* the singular particular propositions are true. Since there cannot be an infinite regress in the first case (cf. *Posterior Analytics*, 72b5–25), it appears that in order to know the truth of any universal proposition we must first establish the truth of at least some universal propositions by investigating the truth or falsity of all their particular instances, and this may well be difficult or impossible.

There are a variety of possible responses to this predicament. One is, in effect, to admit defeat. In the words of the sixteenth-century Portuguese sceptic, Francisco Sanches (addressing an imagined scholastic opponent):

> You admit that there is no science of individuals, because they are infinite. But species are nothing or at least only a certain kind of imagination: only individuals exist, only they are perceived, from them only is science to be had, from them it must be taken; if not, show me these your universals in nature. You will put them in the particulars themselves. I however see nothing universal in them: everything is particular.
>
> (Sanches [1581], p. 126).

Here the connection between nominalism and scepicism is very clear. Another approach is no longer to require certainty, but to be prepared to be satisfied with probabilities. This retreat to probability is the way taken by nearly all modern writers on inductive logic. Yet another possibility is to suppose that in certain circumstances one need investigate only a limited number of particular cases in order to establish firmly the truth of a universal generalisation. One can find a very clear description of this kind of approach in the writings of such sixteenth-century Paduan Aristotelians as Ludovico Buccaferrea (Risse [1964], Vol. I, p. 217) and Jacopo Zabarella. Zabarella held that there is a particular kind of induction, demonstrative induction, which enables us to infer universal conclusions with certainty from an incomplete survey of the particulars. When using this kind of induction 'we do not take all the particulars: for when we begin to enumerate some few of them, it is at once apparent that the predicate is essential to them: therefore leaving aside the enumeration of the remainder, we infer the universal . . .' (*De Methodis* III.14, Zabarella [1608], col. 225F).

In another work, *De Regressu*, Zabarella describes demonstrative induction in a similar way:

> 'Demonstrative induction can be carried on in a necessary subject matter and in things which have an essential connection with one another. It does not therefore consider all the particulars, since after certain of them have been examined, our mind immediately discerns the essential connection, and then, disregarding the remaining particulars, at once infers the universal; for it knows that it is necessary how things must be with the remainder.'
>
> (Zabarella [1608], col. 485D).

This kind of demonstrative induction presupposes the existence of necessary connections between things which are capable of being intuitively grasped by our minds. As Zabarella says '*Inductio . . . demonstrativa fit in materia necessaria*', and if there are in reality no necessary connections to be grasped then the kind of demonstrative induction which Zabarella describes is impossible.

With this statement of Zabarella's theory in mind, we may return to examine its ultimate source, the account of induction in *Posterior Analytics*

II.19. As was remarked earlier, Aristotle appears to slide from an account of how we acquire general concepts to an account of how we get to know general truths. The reason he does this is certainly not philosophical incompetence, or mere carelessness. Rather it is that the two processes are extremely closely connected within the general framework of Aristotle's philosophy. In *Posterior Analytics* II.19 Aristotle was not describing the formation of something like a Lockean abstract general idea. For Aristotle what comes into the soul is a real universal *thing*, a form no longer individuated by matter. This is often concealed for the English reader by translators who put words into their translation which have no counterpart in the original, presumably with the aim of making Aristotle's thought more intelligible to modern readers. Tredennick, for example, in the Loeb edition, gives the following translation of 100a14ff:

> Let us re-state what we said just now with insufficient precision. As soon as one individual percept has 'come to a halt' in the soul, this is the first beginning of the presence there of a universal (because although it is the particular that we perceive, the act of perception involves the universal, e.g., 'man' not 'a man, Callias'). Then other 'halts' occur among these proximate universals, until the indivisible genera or ultimate universals are established. E.g. a particular species of animal leads to the genus 'animal', and so on. Clearly then it must be by induction that we acquire knowledge of the primary premises, because this is also the way in which general concepts are conveyed to us by sense-perception.

The word 'concept' in the last sentence has however no equivalent in the Greek text, which refers merely to universals (cf. 100a6). These universals are not mere 'general ideas', of the kind which appear in seventeenth- and eighteenth-century theories of knowledge. They are on the contrary real universal things existing in the intellect (*De Anima*, 417b23), and it is their presence there which makes possible the kind of demonstrative induction described by Zabarella. Once their existence is denied then two things begin to happen which determine much of the distinctive character of modern philosophy. If there are no real universals, then the only things to which universality can intelligibly be ascribed are signs which may be words or mental concepts or ideas, however conceived. Propositions about such signs are still possible objects of human knowledge, but inevitably they appear more and more limited in scope. Plato and Hume are superficially in agreement in regarding the relations of ideas as the true objects of human knowledge. The intellectual distance between them is a consequence of the utterly different meanings which they attached to the word 'idea'. Propositions of this kind are, according to Hume, 'discoverable by the mere operation of thought, without dependence on what is anywhere existent in the universe' (Hume [1748], p. 25). They can therefore tell us nothing whatever about the existence or nature of anything in that universe. Propositions about matters of fact and existence cannot be shown to be true or false by any kind of intellectual intuition, including the kind of intuitive induction described by Zabarella. The only way in which such propositions can even in principle be shown to be true is by ordinary non-intuitive induction, and this appears capable at best of furnishing us with probabilities, and at worst of giving us no reason whatever for accepting the truth of any universal generalisation. The problem of induction emerges as one of the central problems of modern philosophy.

REFERENCES

AARSLEFF, H. [1982]: *From Locke to Saussure.* Minneapolis: Minnesota University Press.

ARNAULD, A. and NICOLE, P. [1662]: *La logique ou l'art de penser*, (*ed.*) P. Clair and F. Girbal. Paris: Presses Universitaires de France, 1965.

BACON, F. [1859]: *The Philosophical Works of Francis Bacon*, (*ed.*) R. L. Ellis and J. Spedding. London: Longmans.

BARNES, J. [1975]. *Aristotle's Posterior Analytics.* Oxford: Clarendon Press.

BARNES, J. [1982a]: 'The beliefs of a Pyrrhonist', *Proceedings of the Cambridge Philological Society*, **208**, pp. 1–29.

BARNES, J. [1982b]: 'Medicine, Experience and Logic', in J. Barnes *et al.* (*eds.*), *Science and Speculation.* Cambridge: Cambridge University Press.

BARNES, J. [1983]: 'Ancient Skepticism and Causation', in M. F. Burnyeat (*ed.*) *The Skeptical Tradition.* Berkeley: University of California Press.

BARROW, I. [1734]: *The Usefulness of Mathematical Learning explained and demonstrated.* London.

BERKELEY, G. [1713]: *Three Dialogues between Hylas and Philonous*, in A. A. Luce and T. E. Jessop (*eds.*), *The Philosophical Works of George Berkeley*, vol. II. London: Nelson, 1949.

BLACKBURN, S. [1976]: Review of I. Hacking, *The Emergence of Probability, Philosophy*, **51**, pp. 476–80.

BOYLE, R. [1772]: *The Works of the Honorable Robert Boyle.* London.

BURNYEAT, M. [1982]: 'The Origins of non-deductive inference', in J. Barnes *et al.* (*eds.*) *Science and Speculation.* Cambridge: Cambridge University Press.

BURNYEAT, M. [1984]: 'The Sceptic in his place and time', in R. Rorty *et al.* (*eds.*) *Philosophy in History.* Cambridge: Cambridge University Press.

DESCARTES, R. [1970]: *Philosophical Letters.* (*ed.*) A. Kenny. Oxford: Oxford University Press.

ENGBERG-PEDERSEN, T. [1980]: 'More on Aristotelian Epagoge', *Phronesis*, **24**, pp. 301–19.

GARBER, D. and ZABELL, S. [1979]: Review of I. Hacking, *The Emergence of Probability, Archives for the History of Exact Sciences*, **21**, pp. 33–53.

GASSENDI, P. [1658]: *Opera Omnia.* Lyon.

GASSENDI, P. [1972]: *The Selected Works of Pierre Gassendi.* (*ed.*) C. B. Brush. New York: Johnson Reprint Corporation.

GLANVILL, J. [1676]: *Essays on several important Subjects.* London.

GUTHRIE, W. K. C. [1981]: *A History of Greek Philosophy*, vol. VI. Cambridge: Cambridge University Press.

HACKING, I. [1975]: *The Emergence of Probability.* Cambridge: Cambridge University Press.

HACKING, I. [1981]: 'From the Emergence of Probability to the Erosion of Determinism', in J. Hintikka, D. Gruender and E. Agazzi (*eds.*), *Probabilistic Thinking, Thermodynamics and the Interaction of the History and the Philosophy of Science.* Dordrecht: Reidel.

HAMLYN, D. W. [1976]: 'Aristotelian Epagoge', *Phronesis*, **21**, pp. 167–84.

HOBBES, T. [1839]: *Opera Philosophica quae Latine scripsit Omnia*, (*ed.*) W. Molesworth. London: Bohn.

HOOKE, R. [1705]: *The Posthumous Works of Robert Hooke.* London.

HUME, D. [1739]. *A Treatise of Human Nature*, (*ed.*) L. A. Selby-Bigge (2nd edition, revised P. H. NIDDITCH). Oxford: Clarendon Press, 1978.

HUME, D. [1745]: *A Letter from a Gentleman to his friend in Edinburgh*, (*eds.*) E. C. Mossner and J. V. Price. Edinburgh: Edinburgh University Press, 1967.

HUME, D. [1748]: *Philosophical Essays concerning Human Understanding*, (*ed.*) by L. A. Selby-Bigge as *Enquiries concerning Human Understanding and concerning the Principles of Morals* (3rd edition, revised P. H. Nidditch). Oxford: Clarendon Press, 1975.

HUME, D. [1751]: *An Enquiry concerning the Principles of Morals*. Details of edition as in Hume 1748.

KAHN, C. H. [1981]: 'The Role of *nous* in the Cognition of First Principles in the Posterior Analytics', in E. Berti (*ed.*), *Aristotle on Science: the Posterior Analytics*. Padua: Editrice Antenore.

KEYNES, J. M. [1921]: *A Treatise on Probability*, in *The Collected Writings of John Maynard Keynes*, Vol. VIII. London: Macmillan, 1973.

LAUDAN, L. (1981]: *Science and Hypothesis*. Dordrecht: Reidel.

LEIBNIZ, G. W. VON [1765]: *New Essays on Human Understanding*, trans. P. Remnant and J. Bennett. Cambridge: Cambridge University Press, 1981.

LEIBNIZ, G. W. [1903]: *Opuscules et fragments inédits de Leibniz*, (*ed.*) L. Couturat. Paris: Presses Universitaires de France.

LEIBNIZ, G. W. [1969]: *Philosophical Papers and Letters*, (*ed.*) L. Loemker. Dordrecht: Reidel.

LOCKE, J. [1823]: *The Works of John Locke*. London.

MILTON, J. R. [1981]: 'John Locke and the Nominalist Tradition'. in R. Brandt (*ed.*) *John Locke*. Berlin: De Gruyter.

OCKHAM, W. [1980]: *Opera Theologica*, vol. XI, ed. J. C. Wey. St Bonaventure N.Y.: St Bonaventure University Press.

PARKER, S. [1666]: *A Free and Impartial Censure of the Platonic Philosophy*. Oxford.

POPPER, K. R. [1959]: *The Logic of Scientific Discovery*. London: Hutchinson.

POPPER, K. R. [1963]: *Conjectures and Refutations*. London: Routledge and Kegan Paul.

POPPER, K. R. [1972]: *Objective Knowledge*. Oxford: Clarendon Press.

RISSE, W. [1964]: *Die Logik der Neuzeit*. Stuttgart: Friedrich Fromann.

ROSS, W. D. [1949]: *Aristotle's Prior and Posterior Analytics*. Oxford: Clarendon Press.

SANCHES, F. [1581]: *Quod nihil scitur*, (*ed.*) S. Rabade *et al.* Madrid: Consejo Superior de Investigaciones Cientificas, 1984.

SEDLEY, D. [1973]: 'Epicurus, On Nature, Book XXVIII', *Cronache Ercolonesi*, **3**, 1973, pp. 5–83.

SHAPIRO, B. [1983]: *Probability and Certainty in seventeenth-century England*. Princeton: Princeton University Press.

SMIGLECKI, M. [1638]: *Logica*. Oxford.

STILLINGFLEET, E. [1710]: *The Works of Dr Edward Stillingfleet*. London.

STOUGH, C. [1969]: *Greek Scepticism*. Berkeley: University of California Press.

STOUGH, C. [1984]: 'Sextus Empiricus on Non-Assertion', *Phronesis*, **29**, pp. 137–64.

UPTON, T: V. [1981]: 'A note on Aristotelian *epagoge*', *Phronesis*, **26**, pp. 172–6.

VAN LEEUWEN, H. G. [1970]: *The Problem of Certainty in English Thought, 1630–1690*. The Hague: Martinus Nijhoff.

VENN, J. [1889]: *The Principles of Empirical or Inductive Logic*. London: Macmillan.

WALLIS, J. [1687]: *Institutio Logica*. Oxford.

WILKINS, J. [1675]: *The Principles and Duties of National Religion*. London.

WOLFSON, H. [1976]: *The Philosophy of the Kalam*. Cambridge, Mass.: Harvard University Press.

WOLTER, A. [1962]: *Duns Scotus: Philosophical Writings*. Edinburgh: Nelson.

ZABARELLA, J. [1608]: *Opera Logica*. Frankfurt.

Discussions

A RETRACTION OF 'A GEDANKEN EXPERIMENT TO MEASURE THE ONE-WAY VELOCITY OF LIGHT'

My (Nissim-Sabat, 1984) paper proposing a method for measuring the speed of light on a one-way trip is wrong and I am retracting it. Briefly, I had proposed that if one had two identical clocks at A and B respectively one could measure the speed of light from A to B: the fact that the two clocks could not be synchronized would be remedied by the interchange of the positions of the two clocks. Winnie [1970] has shown that if the two clocks are transported in exactly the same way (*i.e.*, with identical velocities as measured by round trip methods), then the two clocks would show exactly the same time lapse (including relativistic time dilation) for their respective trips. Thus I had argued that any inherent time difference between the two clocks having been preserved, its effect would have been cancelled out after the interchange. Yet, following Winnie's reasoning, one can show that all observations made this way would lead experimenters to conclude that the velocity of light from A to B is the same as that from B to A, regardless of whether this is indeed the case. One objection that one can make to my method is to point out that I had assumed that laboratory bound observers would necessarily agree with the moving clocks that the trip from A to B took the same time as that from B to A.

I was led to the above reconsiderations by the objections of an anonymous referee for the *American Journal of Physics* to another synchronization method I had presented.

CHARLES NISSIM-SABAT
Northeastern Illinois University

REFERENCES

NISSIM-SABAT, C. [1984]: 'A Gedanken Experiment to Measure the One-Way Velocity of Light'. *Brit. J. Phil. Sci.*, 35, 62–64.

WINNIE, J. [1970]: 'Special Relativity Without One-Way Assumptions'. *Phil. Sci.*, 37, 81–99 and 223–238.

PUTNAM ON INCOMMENSURABILITY

(1) In Putnam [1981], p. 114 Putnam asserts that 'both of the two most influential philosophies of science of the Twentieth Century . . . are self

refuting'. The philosophies he has in mind are logical positivism and the historical approach. I shall discuss an idea that belongs to the latter viz. incommensurability and I shall show that while the idea may have unusual consequences, self refutation is not one of them.

(2) According to Putnam 'the incommensurability thesis is the thesis that terms used in another culture, say, the term "temperature" as used by a seventeenth century scientist, cannot be equated in meaning and reference with any terms or expressions *we* possess' (114). I shall call the incommensurability thesis as defined in this statement T.

To refute T Putnam points out,

(A), that 'if [T] were really true, then we could not translate other languages—or even past stages of our language—at all' (114), adds,
(B), that 'if Feyerabend . . . were right, then members of other cultures, including seventeenth century scientists would be conceptualisable by us only as animals producing responses to stimuli' and concludes,
(C), 'To tell us that Galileo has "incommensurable" notions ***and then to go on and to describe them at length*** is totally incoherent' (114f—Putnam's italics).

(3) A, B and C rest on the following two assumptions: understanding foreign concepts (foreign cultures)

[i] requires translation and
[ii] a successful translation does not change the translating language.

Neither [i] nor [ii] is correct. We can learn a language or a culture from scratch, as a child learns them, without detour through our native tongue (linguists, historians and anthropologists, having realised the advantages of such a procedure now prefer field studies to the reports of bilingual informants). And we can change our native tongue so that it becomes capable of expressing alien notions (successful translations always change the medium in which they occur: the only languages satisfying [ii] are formal languages and the languages of tourists).

Modern lexica exploit both possibilities. Instead of the semantic equations that formed the basis of older dictionaries they employ research articles of an open and speculative nature. (An example are the introduction and the major research articles of Snell [1971].) Analogies, metaphors, negative characterizations, bits and pieces of cultural history are used to present a new semantic landscape with new concepts and new connections between them. Historians of science proceed in a similar way, but more systematically. Explaining, say, the notion of 'impetus' in 16th and 17th century science they first teach their readers the physics, metaphysics, technology, even the theology of the time, *i.e.* they again introduce a new and initially unfamiliar semantic landscape and then show where impetus is located in it. Examples are found in the work of Pierre Duhem, Anneliese Maier, Marshall Clagett, Hans Blumenberg and, for other concepts, the work of Ludwik Fleck and Thomas Kuhn.

Translating a language into another language is in many ways like constructing a scientific theory; in both cases we must find concepts that fit the 'language of the phenomena'. In the natural sciences the phenomena are those of inanimate nature. Nobody doubts that it is difficult to give a general account of them, that we may have to revise the terms with which we started and that we may have to revise them further when new phenomena appear. In the case of translation the phenomena are the ideas implicit in another language. These ideas developed in different and often unknown geographical surroundings, under different and often unknown social circumstances and they went through numerous intended and unintended changes (influence of further languages, deterioration, poetic licence, etc.). Putnam's [ii] assumes that every language contains everything that is needed for dealing with all these eventualities. To use an example, it makes the rather unlikely assumption that modern Suaheli is already adapted to the language of the Eskimo and, therefore, to Eskimo history. There are only two ways in which such an assumption could succeed: apriorism, or preestablished harmony. Being an empiricist I reject both.

(4) According to Putnam T makes it impossible to explain foreign (primitive, technical, ancient) concepts in English—this is the content of C. He is right in one sense, wrong in another. It is indeed impossible, and trivially so, to formulate ideas in a language not fit to receive them. But the criteria which identify a natural language do not exclude change. English does not cease to be English when new words are introduced or old words given a new sense. Every philologist, anthropologist, sociologist who presents an archaic (primitive, exotic, etc.) world view, every popular science writer who wants to explain unusual scientific ideas in ordinary English, every surrealist, dadaist, teller of fairy tales, ghost stories, science fiction novels, every translator of the poetry of different ages and nations knows how first to construct, out of English *words*, an English *sounding* model of the pattern of usage he needs and then to adopt the pattern and to 'speak' it. A rather trivial example is Evans-Pritchard's explanation of the Azande word *mbisimo* designating the ability of their poison oracle to see far-off things. In his [1975], p. 151 Evans-Pritchard 'translates' *mbisimo* as 'soul'. He adds that it is not soul in our sense, implying life and consciousness but a collection of public or 'objective' events. The addition modifies the use of the word 'soul' and makes it more suitable for expressing what the Azande had in mind. Why 'soul' and not another word? 'Because the notion this word expresses in our own culture is nearer to the Zande notion of *mbisimo* of persons than any other English word'—*i.e.* because of an *analogy* between the English soul and the Azande *mbisimo*. The analogy is important for it smoothes the transition from the original to the new sense; we feel that despite the change of meaning we are still speaking the same language. Now if a conceptual change like the one just described does not go through a metalanguage but stays in the language itself (in which case we would speak of changing the properties of things rather than the usage of words) and if it is not only a single term but an entire conceptual system that is being

received then we have the situation alluded to in (C), but defused, for the English with which we start is not the English with which we conclude our explanation.

(5) Azande ideas already exist in a spoken language and English notions were changed to accommodate them. There are other cases where linguistic change introduces a novel and as yet unexpressed point of view. The history of science contains many examples of this kind. I shall explain the matter by taking an example from the history of ideas.

In *Iliad* 9, 225ff Odysseus tries to get Achilles back into the battle against the Trojans. Achilles resists. 'Equal fate' he replies 'befalls the negligent and the valiant fighter; equal honor goes to the worthless and to the virtuous' (318f). He seems to say that honor and the appearance of honor are two different things.

The archaic notion of honor did not allow for such a distinction. Honor, as understood in the epic was an aggregate consisting of partly individual partly collective actions and events. Some of the elements of the aggregate were: the position (of the individual possessing or lacking honor) in battle, at the assembly, during internal dissension; his place at public ceremonies; the spoils and gifts he received when a battle was finished and, naturally, his behavior on all these occasions. Honor was present when (most of) the elements of the aggregate were present, absent otherwise (cf. *Il.* 12, 310ff—Sarpedon's speech).

Achilles introduces a different point of view. He was offended by Agamemnon who took his gifts. The offence created a conflict between the individual and the collective ingredients of honor. The Greeks who appeal to Achilles, Odysseus among them, illustrate the customary resolution of the conflict: Achilles' gifts have been returned, more gifts have been promised, harmony has returned to the aggregate, honor has been restored (519, 526, 602f). So far we are squarely within tradition. Achilles moves away from it. Pushed along by his lasting anger he perceives an equally lasting imbalance between personal worth and social rewards. What he has in mind not only differs from the traditional aggregate, it is not even an aggregate for there is no set of events that guarantees the presence of honor as he now sees it. Using Putnam's terminology we can say that Achilles' idea of honor is 'incommensurable' with the traditional idea. And indeed, given the epic background the short excerpt I quoted from Achilles' speech sounds as nonsensical as the statement 'equal time needs the fast and the slow to reach the goal'. Yet Achilles introduces his idea in the very same language that seems to exclude it. How is that possible?

It is possible because, like Evans-Pritchard, Achilles can change *concepts* while retaining the associated *words*. And he can change concepts without ceasing to speak Greek because concepts are ambiguous, elastic, capable of reinterpretation, extrapolation, restriction, or, to use a term from the psychology of perception, concepts like percepts obey figure-ground relations.

For example, the tension between the individual and the collective elements of honor that was caused by Agamemnon's deed can be seen in at least two ways, as involving ingredients of equal weight, or as a conflict between fundamental and more peripheral elements. Tradition accepted the first view or, rather, there was no question of a conscious acceptance—people *simply acted* that way: 'With gifts promised go forth. The Achaians will honor you as they would an immortal!' (602f). Achilles driven by his anger magnifies the tension so that it turns from a transitory disturbance into a cosmic rift (figure-ground relations often change as a result of strong emotions; this is the principle of the Rohrschach-test). The extrapolation does not void his speech of meaning because there exist analogies for what he is trying to express. Divine knowledge and human knowledge, divine power and human power, human intention and human speech (an example used by Achilles himself: 312f) are opposed to each other as Achilles opposes personal honor and its collective manifestations. Guided by the analogies his audience is drawn into the second way of seeing the tension and so discovers, as Achilles did, a new side of honor and of archaic morality. The new side is not as well defined as the archaic notion—it is more a foreboding than a concept—but the foreboding produces new ways of speaking and thus, eventually, clear new concepts (the concepts of some Presocratic philosophers are endpoints of this line of development). Taking the unchanged traditional concepts as a measure of sense we are of course forced to say that Achilles speaks nonsense (cf. Parry [1956] for this point; cf. also Feyerabend [1975], 267). But measures of sense are not rigid and unambiguous and their changes are not so unfamiliar as to prevent the listeners from grasping what Achilles has in mind. Speaking a language or explaining a situation, after all, means both *following* rules and *changing* them; it is an almost inextricable web of logical and rhetorical moves.

From what has just been said it also follows that speaking a language goes through stages where speaking indeed amounts to merely 'making noises' (Putnam [1981], 122). For Putnam this is a criticism of the views he ascribes to Kuhn and myself. For me it is a sign that Putnam is unaware of the many ways in which language can be used. Little children learn a language by attending to noises which, being repeated in suitable surroundings, gradually assume meaning. Commenting on the explanations which his father gave him on logical matters Mill writes (Lerner [1965], 21): 'The explanations did not make the matter clear to me at the time; but they were not therefore useless; they remained as a nucleus for my observations and reflections to crystallize upon; the import of his general remarks being interpreted to me, by the particular instances which came under my notice *afterwards*.' Saint Augustine advised parsons to teach the formulae of the faith by rote adding that their sense would emerge as a result of prolonged use within a rich, eventful and pious life. Theoretical physicists often play around with formulae which do not yet make any sense to them until a lucky combination makes everything fall into place (in the case of the quantum theory we are still waiting for this lucky combination). And Achilles, by his way of talking, created new speech habits which eventually gave rise to new and more abstract conceptions of honor, virtue and being. Thus using words as mere noises has an important function even within the most advanced stages of speaking a language (cf. Feyerabend [1975], 270).

One of the scientists who was aware of the complex nature of explanatory talk and who used its elements with superb skill was Galileo. Like Achilles Galileo gave new meanings to old and familiar words; like Achilles he presented his results as parts of a framework that was shared and understood by all (I am now speaking of his change of basic kinematic and dynamical notions); but unlike Achilles he knew what he was doing and he tried to conceal the conceptual changes he needed to guarantee the validity of his arguments. Chapters 6 and 7 of my [1975] contain examples of his art. Taken together with what I have said up till now these examples show how it is possible to assert, without becoming incoherent, that the Galilean notions are 'incommensurable' with our own 'and then to go on and to describe them at length'.

(6) They also solve Putnam's conundrum about the relation between relativity and classical mechanics. If T is correct, says Putnam, then the sense of statements that occur in a test of either relativity or classical mechanics cannot be 'independent of the choice between Newtonian and Einsteinian theory'. Moreover, it is then impossible to find equations of meaning between 'any word in . . . Newtonian theory [and] any word in . . . general relativi[ty]' (116). He infers that there is no way to compare the two theories. The inference is again mistaken. As I mentioned in section 3 linguists long ago ceased using equations of meaning to explain new and unfamiliar ideas while scientists always emphasized the novelty of their discoveries and of the concepts used in their formulation. This does not stop them from comparing theories, however. Thus the relativist can say that the classical formulae, ***properly interpreted*** (*i.e.* interpreted in the relativistic manner) are successful, but not as successful as the full relativistic apparatus. He can argue like a psychiatrist who, talking to a patient who believes in demons (Newton) adopts his, the patient's manner of speaking without accepting its demonic (Newtonian) implications (this does not exclude the the possibility that the patient will one fine day turn around and convince him of the existence of demons). Or he may teach relativity to the classicist like a foreign language and invite him to judge its virtues from within ('having learned Spanish to perfection and having read Borges and Vargas Llosa, would you not rather write stories in Spanish than in German?') There exist many other ways in which the Newtonian and the relativist can and do converse. I have explained them in papers I wrote in 1965, some of them in direct response to Putnam's criticisms of that time: Feyerabend [1981], Vol. i, chapter 6, sections 5ff and Vol. ii, chapter 8, section 9ff and appendix. This finishes my response to A, B and C.

(7) The arguments of the preceding sections were based on T which is Putnam's version of incommensurability. But Putnam's version is not the version I introduced when examining the relation between comprehensive theories such as Newton's mechanics and relativity or Aristotelian Physics and the new mechanics of Galileo and Newton (cf. Feyerabend [1975] 268ff and [1981], Vol. i, chapter 4, section 5). There are two differences. First, incommensurability as understood by me is a rare event. It occurs only when the conditions of meaningfulness for the descriptive terms of one language (theory, point of view) do not permit the use of the descriptive terms of another language (theory, point of view); mere difference of meanings does not yet lead to incommensurability in my sense. Secondly, incommensurable languages (theories, points of view) are not completely disconnected—there exists a subtle and interesting relation between their conditions of meaningfulness. In my [1975] I explained this relation in the case of Homeric Commonsense vs. the language aimed at by the early Greek philosophers. In [1981] Vol. i, chapter 4 I explained it in the case of Aristotle-Newton. I should add that incommensurability is a difficulty for philosophers, not for scientists. Philosophers insist on stability of meaning throughout an argument while scientists, being aware that 'speaking a language or explaining a situation means both *following* rules and *changing* them' (see section 5 of this paper) are experts in the art of arguing across lines some philosophers regard as insuperable boundaries of discourse.

PAUL FEYERABEND
University of California
at Berkeley

REFERENCES

EVANS-PRITCHARD, E. E. [1976]: *Witchcraft, Oracles and Magic Among the Azande*, abridged edition, Oxford, 1976.

FEYERABEND, PAUL K. [1975]: *Against Method*, London, 1975.

FEYERABEND, PAUL K. [1981]: *Philosophical Papers*, 2 Vols., Cambridge, 1981.

LERNER, MAX [1965]: *Essential Works of John Stuart Mill*, New York, 1965.

PARRY, A. [1956]: 'The Language of Achilles', *Trans. & Proc. Amer. Philos. Assoc.*, Vol. 87, 1956.

PUTNAM, HILARY [1981]: *Reason, Truth and History*, Cambridge, 1981.

SNELL, BRUNO [1971]: *Lexikon des Fruehgrechischen Epos*, Goettingen, 1971.

PAPINEAU ON CAUSAL ASYMMETRY

David Papineau in 'Causal Asymmetry' attempts to provide an account of causal asymmetry which does not depend upon the assumption that causes precede their effects in time [1985]. In place of a temporal-priority view, he introduces a probabilistic analysis. This analysis takes two forms, one appropriate to determinism and the other appropriate to indeterminism. I will argue that Papineau's account is subject to a series of counterexamples.

According to Papineau, under determinism, causal asymmetry 'derives from the fact that the background conditions together with which causes determine their effects are independent of each other, whereas the same does not hold of the background conditions with which effects 'determine' their causes' ([1985], p. 280). The pair of equivalences, $S\&X \vee A \leftrightarrow C$ and $S8M \vee N \leftrightarrow Y$, associated with a common cause S (smoking) of joint effects C (cancer) and Y (yellow fingers), will be such that the background conditions $(X \vee A)$ and $(M \vee N)$, are probabilistically independent. However, the pair of equivalences, $C\&-A \vee T \leftrightarrow S$ and $C\&-S \mathrm{V} R \leftrightarrow A$, associated with a common effect C of alternative causes S and A (asbestos working) will be such that the 'background conditions,' $(-A \vee T)$ and $(-S \vee R)$, are *not* independent. Similarly, under indeterminism, in the case of joint effects, if we treat X and M in these equivalences as the indeterministic stages needed for S to 'go through' to C and Y, respectively, then the background conditions $(X \vee A)$ and $(M \vee N)$ will be independent. Papineau claims that no such independence will be found in the case of 'background conditions' to alternative indeterministic causes of a common effect.

Contrary to Papineau's analysis, there are cases in which a cause C_1 is causally prior to its effect C_2 under determinism, but the relevant background conditions do not exhibit the kind of independence posited by Papineau. Aware that this possibility poses a serious difficulty for his account, Papineau attempts to extend his approach so as to handle such *prima facie* counterexamples. In considering the possibility that A (asbestos working) might be probabilistically associated with $-M$ (finger scrubbing), contrary to the assumption made in the example above of joint effects, he argues that 'in order for there to be such a correlation between asbestos working and finger scrubbing, there would then have to be a common cause of these two factors . . . , and I would argue that the initial conditions in conjunction with which *that* common cause gave rise to its joint effects would then be independent of each other' ([1985], p. 281).

We may generalize from this response. His strategy for dealing with dependent background conditions is to trace back from these conditions to a common cause which operates to generate these background conditions in virtue of further, *independent* background conditions. This strategy for dealing with this class of problem cases can be shown, however, not to be adequate for the following series of cases with dependent background conditions.

First, given the possibility that background conditions, in the case of a common cause, may fail to be independent, we cannot rule out the possibility that the common cause, to which A and $-M$ are traced, *itself* operates in association with dependent background conditions. We might, of course, further extend Papineau's strategy of 'tracing back' until some common cause was discovered which operated in association with independent background factors. This process of 'tracing back,' however far it is extended, will not help us deal with the following, possible causal chain: H is a causal chain (with multiple branches) which extends infinitely into the past (in the case of forward causation), and is such that for any common cause C in H, C operates in conjunction with dependent background conditions. In the case of causal chain with this characteristic, Papineau's analysis, even as extended along the lines he suggests, will not provide a basis for assigning causal direction. Further, given his admission that independence will sometimes fail in the case of background conditions of a common cause, the possibility of causal chains of this type arises.

In addition to such chains there are other types of cases involving common causes with dependent background conditions which are not amenable to Papineau's analysis. Background conditions to a common cause may fail to be independent even if they themselves are not the result of a common cause. Consider, for example, background conditions to a common cause which are not independent in that they are identical. Suppose that U has two effects, X and W, in virtue of the same set of background conditions as specified in the following equivalences:

$$U\&Z \vee O \leftrightarrow X$$

$$U\&Z \vee O \leftrightarrow W$$

U causes two effects in conjunction with the same set of background conditions ($Z \vee O$). We cannot trace back from either factor to a common cause operating in conjunction with additional, independent background conditions. Such a possibility only arises in the case of non-identical background conditions.

Even, however, when background conditions to a common cause are non-identical, independence may fail without it being true that this dependence can be accounted for by reference to a common cause. Background conditions to a common cause may fail to be independent in virtue of some partial overlap of factors. Again, suppose that U has two effects, X and W, but in this case the relevant background exhibit only partial identity:

$$U\&H \vee G \leftrightarrow X$$

$$U\&M \vee G \leftrightarrow W$$

The background conditions, ($H \vee G$) and ($M \vee G$), are non-identical, but may not be independent since G appears in both conditions. Since we can assume that the distinct factors, H and M, do not result from a common cause, there is no room for applying Papineau's strategy in order to handle problem cases of this type.

In turn, background conditions to a common cause may fail to be independent even if they are distinct, but not the result of a common cause. Suppose that the following equivalences hold:

$$T\&M \vee N \leftrightarrow D$$

$$T\&G \vee H \leftrightarrow Y$$

($M \vee N$) and ($G \vee H$) may not be independent if G and M or H and N, for instance, are directly causally connected. If G causes M, for example, then ($M \vee N$) and ($G \vee H$) may fail Papineau's independence criterion. But in such cases there will not be a common cause to which this lack of independence can be traced. This will be so even if G is caused by some event C. Although C will then count as a cause of M, given that G causes M and the transitivity of causation, it will not follow that C is a common cause of G and M in virtue of independent background conditions. C causes M partly in virtue of the background conditions which are relevant to C's causing G. Hence, given this partial overlap, there will not exist appropriate independent background conditions. Indeed, C is not a *common cause* of C and M, but rather causes M indirectly by way of causing G. A similar failure of independence would arise which is not traceable to a common cause if G were to *prevent* M from occurring rather than causing M. In that case, G and M would be negatively correlated. But again, we need not assume that G and M or $-M$, are indirectly linked by way of a common cause. Hence, Papineau's strategy cannot be applied.

In considering the other side of the coin, there are cases involving a joint effect in which the 'inverted' background conditions associated with the alternative causes *are* probabilistically independent, contrary to Papineau's account. Probabilistic independence among 'inverted' background conditions will arise under certain types of causal preemption. In order to establish this point I will modify the example offered by Papineau in which S (smoking) and A (asbestos) are alternative causes of C (cancer). The relevant equivalences are as follows:

$$C\&-A \vee T \quad S$$

$$C\&-S \vee R \quad A$$

As Papineau presents this case, the 'inverted' background conditions, ($-A \vee T$) and ($-S \vee R$), fail to be independent because $-A$ and R are negatively correlated as are $-S$ and T. However, this example can be changed in such a way that S and A remain alternative causes of C, but the negative correlation disappears, and ($-A \vee T$) and ($-S \vee R$) turn out to be probabilistically independent.

The 'inverted' background conditions will be probabilistically independent just in case the following equalities hold:

(a) $P(-A \vee T) = P(-A \vee T)/(-S \vee R)$

(b) $P(-S \vee R) = P(-S \vee R)/(-A \vee T)$

Now if we assume with Papineau that R and T are independent, and that S and A are independent, then (a) and (b) will be true if R and A are independent, and S and T are independent. R and A will be independent if the following hold:

(c) $P(R/A) = P(R)$

(d) $P(A/R) = P(A)$

S and R will be independent if similar equalities hold with 'S' and 'T' substituted for 'R' and 'A', respectively. It can be shown that (c) and (d) may hold, even if S and A are alternative causes of C, if R has alternative causes A and B, which satisfy certain conditions. These conditions will include the fact that when A occurs B is prevented from occurring and hence, from causing R, and if A were to fail to occur, B would occur. Let us then suppose that

(e) $P(R/A) = 1$

(f) $(R/B) = 1$

(g) $P(B/-A) = 1$

It follows that $P(R/-A) = 1$, given (f) and (g). Given that $P(R/-A) = 1$ and (e), we have $P(R/A \vee -A) = 1$ from which it can be concluded that $P(R) = 1$. Hence, $P(R/A) = P(R)$. In turn, suppose that $P(A) = P(B) = 1/2$, and that $P(A \vee B) = 1$. Suppose further that if R occurs it is equally likely to have been caused by A as by B, that R occurs only if caused, and that A and B are the only possible alternative causes of R. Given the latter it follows that $P(A/R) = 1/2$. Hence, $P(A/R) = P(A) = 1/2$. Under these conditions, (e) and (d) are satisfied and R and A are probabilistically independent. Similar assumptions can be made in the case of T and S in order to establish their probabilistic independence. In a situation satisfying these constraints, the 'inverted' background condition, $(-A \vee T)$ and $(-S \vee R)$, will be probabilistically independent, contrary to Papineau's assumption.

Papineau, as we have seen, contends that if S is an indeterministic cause of C, then equivalences of the following form will hold, where X and M are indeterministic stages leading to C and Y:

$$S\&X \vee A \leftrightarrow C$$

$$S\&M \vee N \leftrightarrow Y$$

In addition, and most importantly, X and M will be probabilistically independent of each other. As I will now show, however, this last condition, does not hold in all cases. Consider the following type of case: X and M consist of indeterministic chains of events leading respectively from S to C and from S to Y. Such chains may in fact partially overlap, *i.e.*, share in common some events. For example, both S and M may include as an intermediary stage D. We may also suppose that S and M share no other stages in common so that D is the only stage at which these processes overlap. In cases of this type, since there is partial overlap between the processes which link S with C and S with Y, these processes may fail to be probabilistically independent. Finally, modifying Papineau's view to require only that some parts of each process be independent will not help since such partial independence will also arise in the case of a common effect of two alternative causes.

Finally, Papineau claims that any attempt to construct 'inverted' chains of indeterministic stages running 'back' from a common effect to alternative causes will fail to reveal independent 'intervening' chains. For example, the stages required to 'get back' from cancer, the common effect, to smoking, one of the alternative causes, will include the *absence* of the corresponding stages leading back from cancer to asbestos, the other alternative cause. In other words, if S_1 are the stages leading 'back' from C to S, S_2 the stages leading 'back' from C to A, then S_1 and S_2 will be negatively correlated. His argument for this claim seems to be that the absence of the corresponding stages leading to the alternative cause is required 'to ensure the cancer did come from smoking and not from asbestos' ([1985], p. 284). Now this latter claim is ambiguous between the following two possibilities: if both S_1 and S_2 occurred then causal efficacy could not be attributed to S or A, and if both S_1 and S_2 occurred, then causal efficacy could not be attributed to S *rather than* A or to A *rather than* S. Although the second claim may be right, the first is not persuasive. Certainly, we can imagine cases in which both stages occur, and in which the effect occurs. In such cases, we might well treat *both* S and A as causally efficacious in the production of C. Given that possibility, we can further imagine instances in which the occurrence of the alternative stages leading to C are probabilistically independent of each other, contrary to Papineau's claim. For example, S and A might be the initial stages of two indeterministic quantum mechanical processes involving the emission of particles and C the effect of such processes on a Geiger counter. When S_1 and S_2, the stages linking S to C and A to C respectively, occur together, C is caused by both S and A. In turn, S_1 and S_2 will be probabilistically independent of each other.

These various problem cases raise serious doubts about the central claims of Papineau's paper: that background conditions to a common cause are independent under determinism, but that no such independence exists among 'background conditions' to a common effect, and that the intervening stages of a common cause are independent under indeterminism, but that no such independence exists among the 'intervening' stages of a common effect under indeterminism. Such doubts will naturally transfer to the account of causal asymmetry which depends upon the truth of these two claims.

In closing, I would like to outline in rough form an alternative view of causal asymmetry which I have developed elsewhere (Ehring [1982]). Causal asymmetry is determined by the 'circumstantial' character of the causal relation. For any two events *c* and *e* which are causally connected, there is an event f upon which this causal connection depends, *i.e.*, roughly, if f had not occurred, then c and e would not have been causally connected. For example, the presence of oxygen is a condition of a causal connection between striking a match and its lighting. Without oxygen, these two events would not have been causally related. Causal asymmetry is constituted by the fact that this conditioning event f will normally be causally connected to e, but not to c. In other words, this account has it that a cause is a member of a set of events that possess, within that set, a certain causal independence, whereas the effect event is not causally independent of any of these events. Simplifying even more, we may say, that since every effect has more than one cause, causal priority may be read from the causal independence of the multiple causes. Although this account bears certain similarities to that of Papineau, there are at least two important differences. First, the notion of 'causal connection' is left open. As a consequence, the notion of the 'independence' of causes will not necessarily be cashed in by way of probabilistic independence. Second, the causes which display independence are not Papineau's 'background conditions' to a common cause, but the joint causes of a common effect. In turn, joint effects e_1 and e_2 of a common cause c, although perhaps causally independent, will fail to satisfy the requirement that the causal connection of c to e_1 (or e_2) depend upon e_2 (or e_1).[1]

DOUGLAS EHRING
Southern Methodist University

REFERENCES

PAPINEAU, D. [1985]: 'Causal Asymmetry', *British Journal for the Philosophy of Science*, **36**, pp. 273–89.
EHRING, D. [1982]: 'Causal Asymmetry,' *The Journal of Philosophy*, **79**, pp. 761–74.

INTERACTIONISM AND EVOLUTION: A CRITIQUE OF POPPER

1 *Introduction*
2 *Non-physical Entities and the Natural Sciences*
3 *Popper versus Natural Sciences*

1 INTRODUCTION

Popper has written an argument for interactionism (Popper and Eccles

---

[1] I am grateful to the comments of an anonymous referee on an earlier draft of this paper.

[1977]). He supposes that non-physical entities may influence the physical world. He is convinced that the physical world is not causally closed. Popper's rejection of the concept of a causally closed physical world is based on Darwinism ([1977], pp. 72–74). Popper reasons as follows:

(*a*) mental experiences have emerged during evolution;
(*b*) Darwinism explains the emergence of things only if they make a difference ([1977], p. 88);
(*c*) this means that mental experiences must have an influence on the physical world: the physical world is not causally closed.

The theory of evolution is generally considered as a part of biology. A branch of biology in which both evolutionary theory and the question whether we have to assume interactionism, or not, takes a prominent place is ethology, the biology of behaviour. Lorenz, one of the founders of ethology, has stated that the merit of this science is a strict observance of the epistemological discipline, long ago developed by natural science in general, and to the application of comparative phylogenetic methods developed by Darwin in particular (Lorenz [1958]). With respect to interactionism, Tinbergen, the other founder of ethology, has stated that he rejects subjectivism. By subjectivism he means '. . . the procedure of replying to the question "What causes this behaviour?" by referring to a subjective experience'. As an example he gives 'The animal attacks because it feels angry' (Tinbergen [1963]).

If Popper's reasoning is correct there exists a contradiction within ethology. In this paper I hope to show that such a contradiction is not present. Since this affects the essence of Popper's argument I think that we can not consider this as a minor criticism. First I want to show that an introduction of interactionism in ethology would not only be in contradiction with statements of both founders of ethology, but also with the history of biology and even with that of natural science as a whole.

2 NON-PHYSICAL ENTITIES AND THE NATURAL SCIENCES

Physics did not develop until non-physical entities were no longer accepted as causations. If we accept spirits, ghosts or Gods as satisfying causations, no further investigations seem to be necessary or useful. However when we do not, phenomena that need to be explained become challenges that stimulate research. Biology also did not develop until non-physical entities were no longer accepted as causations. Biology developed later than physics. Apparently we have more difficulties in considering living organisms as part of a causally closed universe than in the case of lifeless matter. Some people, known as vitalists, felt it to be a kind of blasphemy to consider organisms as complicated configurations of matter. They claimed that certain life processes were caused by a non-physical force, the life force or *vis vitalis*. Such an attitude does not stimulate research. However, it did indirectly. Researchers who were of opinion that processes in living organisms also had to be analysed in a natural scientific way (the mechanists) were stimulated to prove that it is very well possible to understand such processes without the acceptance of any non-physical force. Because the vitalists had to give up one claim after another Von Bertalanffy ([1949], p. 21) has stated that the history of biology is identical to the rebuttal of vitalism.

Ethology also developed only when non-physical entities were no longer accepted as causations. It took until this century before this was the case. Apparently we are very apt to consider behaviour as caused by non-physical entities, *in casu* mental experiences. If we say that an animal attacks because it is angry, further investigations to explain this behaviour are supposed to be useless. Instead of that, it is said in ethology that the animal attacks because its aggression system is activated. Such models are derived from the physical world; they are attempts to describe the laws that determine the behaviour and allow predictions that stimulate further research (see Van Rooijen [1983c]). That we are very apt to consider behaviour as caused by non-physical entities is not only indicated by the late development of ethology but also by recent attempts to introduce mental experiences as causations in ethology (see for instance Van Rooijen [1983a]). An important reason for this tendency will be our private experience that mental events exist and that they seem to be able to influence physical matter. That mental events are assumed not to be causally potent does not mean that they do not exist. Lorenz [1958] used the metaphor of two aspects (one physiological and one psychological) of the same reality (see also Van Rooijen [1985b]). That we tend to see mental events as causes or effects of physiological processes is according to Lorenz ([1958], p. 269) due to the fact that in daily life we are sometimes more familiar with the psychological and sometimes with the physiological side of events out of a chain of events. As an example he mentions that we say that a severe disappointment causes a heart disorder. Physiological processes belong to the objective world. That means that they are observable by our exterior senses and therefore in principle observable by more than one person. Psychological phenomena are observable by introspection and therefore only accessible to one person. Such phenomena belong to the subjective world.

Another reason that we are apt to consider behaviour as caused by non-physical entities may be our spontaneous tendency to believe that also other human beings and the higher animals possess mental experiences. Lorenz ([1958], p. 266) has written that the student of behaviour, while fully aware that only objective, physiological processes are accessible to the methods of scientific research cannot help believing the reality of subjective, psychological processes. 'No normal man is able to avoid this type of "empathy" and it is simply not honest not to confess to it.' A third reason might be that we like to see ourselves as individuals with a free will and this seems more in accordance with a view of the mind as a non-physical entity than as another aspect of physical matter.

3 POPPER VERSUS NATURAL SCIENCES

In the natural sciences it is assumed that, in principle, it is possible to make an organism (for instance a dog) out of lifeless matter (already a fertilized egg cell will do). We may assume that such a dog will have all the mental experiences of a dog-made dog. Evolutionists believe that the experiment of making a dog out of lifeless matter has already been done during evolution. Even nowadays dogs are believed to make new dogs out of food only.

Popper ([1977], p. 5) regards materialism as a mistake because, as he argues, he does not want to believe that men are machines. He considers materialism as prone to undermine humanistic ethics. In his opinion materialists (including epiphenomenalists) must neglect the reality of human suffering and the significance of the fight against unnecessary suffering. One may question whether it is true that materialism implies that men are machines (and if it were true one could question whether it is not a prejudice to exclude this possibility beforehand). It is true that within the natural scientific approach men are not treated differently from machines. However, epiphenomenalism does not imply that men are machines. The physical similarity between me and a configuration of matter that we call a machine is very superficial compared with all the similarities between me and other human beings (and the higher animals). Therefore there are not as many reasons to assume that machines have a psychological dimension, as there are reasons to make this assumption in relation with the higher organisms. This seems to be a good reason to keep the distinction between men (and the higher animals) and machines. In certain cases machines may perhaps elicit a feeling of empathy. However, as argued more extensively elsewhere (Van Rooijen [1985a]) this is a prescientific measure. Because epiphenomenalism does not imply that men are machines it also does not imply a neglect of the significance of the fight against unnecessary human (and animal, see for instance Van Rooijen [1983b]) suffering. Popper ([1977], p. 5) has mentioned that the great materialist philosophers were almost all upholders of human ethics, while their opponents have sometimes been the opposite. One might even question whether ethics that need interactionism in order to consider men and animals as worthy of moral concern are not more primitive than ethics that need epiphenomenalism.

Popper ([1977], p. 74) has remarked that according to epiphenomenalism, mental experiences would be irrelevant. For example, if somebody reads a paper the decisive thing would not be that this influences his opinion and provides him with information, but only that it changes his brain structure. This contradistinction is somewhat confusing because such processes may be considered as parts of models which predict future behaviour. These processes are not necessarily connected with mental experiences (a machine may also be able to store information and to change its 'opinion').

With respect to the mental experiences occurring during such changes in brain structure it must be stressed that epiphenomenalism implies that these experiences are not different from these changes, except in the way we observe them. The changes that result from reading a paper are relevant, no matter whether they are viewed by our external senses or by introspection. Another question is whether mental experiences are relevant in the sense that we may consider them as adaptations.

Not all properties of a structure that has evolved need to have adaptive value within the Darwinian framework. Adaptive features may have secondary properties, which are adaptively neutral, *i.e.*, make no difference. Only when the presence of a character is not a logical consequence of a structure is the term adaptation justified. For instance, the brown colour in the feathers of many species of birds is mostly not a logical consequence of the structure 'feather'. In many cases, these brown feathers are a protection against predators. In these cases it is justified to call this brown colour an adaptation.

The brown colour of the liver is an example of a case in which it is not justified to use the term adaptation (Van der Klaauw [1949], p. 139). This brown colour is the logical consequence of the presence of biliary ducts and blood vessels. Also the colours of species that live in the dark (for instance the colour of fish species that live in caves where no daylight ever enters; these species (*Anostichy jordani*; *Caecobarbuys geertsi*) have even lost the ability to see) are not to be considered as adaptations. In a similar way mental experiences may well be a logical secondary (non-adaptive) feature of a structure, which evolved because of adaptive properties (Van Rooijen [1981]). Because it is not correct that only things that make a difference have emerged during evolution there exist no contradiction between evolution and the assumption of a causally closed universe. Although the assumption of a causally closed universe is not necessarily correct this conclusion implies that Popper's argument for interactionism does not hold.

JEROEN VAN ROOIJEN
Churchillweg 37c
6707 J.B. Wageningen
The Netherlands

REFERENCES

BERTALANFFY, L. VON [1949]: *Das biologische Weltbild I.* A. Francke, A.G. Verlag.

KLAAUW, C. J. VAN DER [1949]: 'De verschillende aspecten van het aanpassingsprobleem.' *Vakblad voor biologen*, **8**, pp. 137–49.

LORENZ, K. [1958]: 'Methods of approach to the problems of behaviour.' In: K. Lorenz (*ed.*), *Studies in animal and human behaviour*, **2**, pp. 246–8, 1971, Methuen.

POPPER, K. R. and ECCLES, J. C. [1977]: *The Self and its Brain.* Springer.

ROOIJEN, J. VAN [1981]: 'Are feelings adaptations? The basis of modern applied animal ethology.' *Applied Animal Ethology*, **7**, pp. 187–9.

ROOIJEN, J. VAN [1983a]: 'Awareness and Griffin's circular reasoning.' *Animal Behaviour*, **31**, pp. 613–14.

ROOIJEN, J. VAN [1983b]: 'Genetic Adaptation and Welfare.' *International Journal for the Study of Animal Problems*, **4**, pp. 191–7.

ROOIJEN, J. VAN [1983c]: 'Preference tests, motivations, models and welfare.' *Applied Animal Ethology*, **11**, pp. 1–6.

ROOIJEN, J. VAN [1985a]: '*Du-Evidenz*, applied ethology and animal welfare.' *British Veterinary Journal*, **141**, pp. 245–8.

ROOIJEN, J. VAN [1985b]: 'The philosophical position of applied ethology; a reply.' *Applied Animal Behaviour Science*, **14**, 378–383.

Brit. J. Phil. Sci. 38 (1987), 93–116 *Printed in Great Britain*

Review Articles

MEYERSON'S 'RELATIVISTIC DEDUCTION': EINSTEIN VERSUS HEGEL*

1 *The Necessity of Irrationality*
2 *The Rationalisation of Nature*
  (*a*) *The Principle of Identity*
  (*b*) *Legal Explanations*
  (*c*) *Causal Explanations and Conservation Laws*
  (*d*) *Causality, Geometry and Panmathematicism*
  (*e*) *Hegel's Heritage*
  (*f*) *Meyerson's Irrationalist Solution*
  (*g*) *Meyerson's Conception of the* a priori
  (*h*) *The Defects of Meyerson's Position and its Reinterpretation*
3 *The Irreversibility of Time*
  Conclusion

1 THE NECESSITY OF IRRATIONALITY

In this book, which was published in 1925 and whose translation into English was long overdue, Meyerson set out to account not only for the development of Relativity Theory but, more generally, for the way in which science as a whole proceeds. In the first four chapters he gives a synopsis of the theses which he had put forward in his monumental 'De l'Explication dans les Sciences' [1921] (by the way, this work ought also to be translated into English). In the latter part of the 'Relativistic Deduction' he tries to show that Relativity provides an even better illustration of these theses than did earlier-classical-theories. According to Meyerson, notwithstanding positivist claims to the contrary, science proceeds in strictly realist fashion by postulating a world of perduring objects. The scientist posits a reality transcending the human mind, then goes on to explain this reality in rational terms. However, in the last and, for the philosopher, in the first, analysis rationalism and realism prove incompatible. For *a priori* reasons, all attempts aimed at the total 'rationalisation' of Nature must, according to Meyerson, ultimately fail. He writes:

> Science is realistic; we know however that if explanation is pushed to its limits it can only end up at acosmism—the destruction of reality. In relativism, precisely because it is a very advanced, a very perfect form of theoretical explanation, these

---

* EMILE MEYERSON [1985]: *The Relativisitc Deduction* (translated from the French by David A. and Mary-Alice Sipfle). D. Reidel Publishing Co. liii+268. (£30.50).

two extremes of existence and non-existence are found to be quite close to each other. This leads to a painful sort of conflict for the physicist. (p. 140)

And

> . . . although we make an effort to discover the rational in the real—to demonstrate that reality is entirely rational or, inversely, that reason alone is capable of constituting reality—as soon as we believe we have more or less attained this goal, any part of reality we have shown to be rational is apt to seem to us not to belong to true reality. There is an element of paradox in our minds here . . . (p. 136)

In these passages Meyerson clearly expresses his view that science rests on a sort of Hegelian contradiction of which the scientist may well be unaware: by trying to explain reality science aims in effect at destroying it; but science also takes account of certain facts which, being given rather than explained, represent irrational elements within all explanatory theories. Meyerson in effect identifies the real with the irrational. In an eloquent passage he tells us that a character in a play strikes us as being authentic, *i.e.*, as being real, only to the extent that he or she has certain inexplicable traits (*cf.* p. 136).

Let us now list some of the irrational aspects which, according to Meyerson, science must retain in order to capture reality:

(1) The qualitative aspects of sensation involved in all scientific observation. Following Leibniz, Meyerson maintains that the transition from brain process to sensation, a transition which we simply must accept if we are to test scientific hypotheses, remains unintelligible to us. This irrational jump from the domain of physics to that of consciousness will always bedevil the mind-body problem (*cf.* pp. 91, 133, 165, 200).

(2) The physical interpretation of mathematical systems, more particularly of space-time geometry. Meyerson regards such interpretations as something superimposed on a system which, in order to be rational, should have remained abstract (*cf.* pp. 134–5).

(3) All factual statements not deduced from *a priori* principles. Meyerson asserts that

> what is revealed by discontinuity, whether of atoms or quanta, is, as in the case of Carnot's principle, the component of reality that cannot be deduced from the fundamental principles of reason. Moreover, since in the theory of relativity the rational appears to be purely spatial—geometrical—in nature, one can say that what prevents us from making everything rational in this case is the properly physical, as opposed to mathematical, component of reality. (*cf.* p. 109)

(4) Entropy and the uni-directionality of time. (See quotation above.)

(5) The discontinuous aspect (quantum jumps) of atomic physics. Summing up, Meyerson writes:

> Obviously, things can never really be made completely rational. To explain irreversibility, an improbable state at the beginning of time must be accepted as given. To explain transitive action there are several givens of the same sort that relativity theory cannot claim to have made rational: on the one hand there is that of spatial

discontinuity (Chapter 13) and, on the other, that of interpretation (Chapter 14). (p. 177)

Note that (1)–(5) form a laundry list whose items have no obvious property in common. In his excellent introduction, which contains both a synopsis of Meyerson's philosophy of science and a detailed discussion of the problem of time-irreversibility, Milič Capek justifiably asks ". . . but does not then his [Meyerson's] 'l'irrationnel' become a more convenient escape-route which is used whenever the facts do not fit his analysis? Is not such an escape route merely a disguised impasse?" (p. l).

In 'De l'Explication dans les Sciences', Chapter XI, Meyerson blamed Hegel for accepting only one kind of irrationality whereas he, Meyerson, could discern many kinds. He then wondered whether there was not after all some denominator common to all these seemingly different sorts of irrationality. A large part of the present article is devoted to answering this question which was left open both in 'De l'Explication' and in the 'Relativistic Deduction'. In order to circumscribe what Meyerson might have meant by 'irrationality' it seems best to start by determining what he essentially understood by a '*rational explanation*'.

2 THE RATIONALISATION OF NATURE

(*a*) *The Principle of Identity*

According to Meyerson, all explanations, whether scientific or commonsensical, spring from one basic tendency of the human mind; namely the tendency to deny diversity and change; or to assert the existence of constants behind the fleeting appearances; or to explain the Many in terms of the One; or to subsume the flux of Becoming under the immutability of Being. The best formulation, in my opinion, is that the human mind inevitably tends to deny diversity and assert sameness or identity both in space and in time. For the human mind only the undifferentiated One is real, everything else is appearance. Note that I have spoken of this principle, as of a tendency inherent in the human mind. If this tendency went unchecked, it would assert the existence of a unique Parmenidean Sphere which is undifferentiated in space and changeless in time; it would thus degrade all phenomena to the rank of mere sensory illusions. Becoming would thus be mere illusion and only Being would be real. The whole of science can be conceived as a sequence of attempts to rescue as much of Parmenides's sphere as is compatible with the phenomena. It is well-known that Greek Atomism postulated the void together with a plurality of particles each of which can be assimilated to a Parmenidean Sphere. Both the atoms and the void are immutable in themselves. The flux of appearances is explained in terms of different configurations of the same particles within the same empty space. Let A and K stand for two consecutive states of affairs which, following Meyerson, will henceforth be referred to as the antecedent and the consequent respectively. Thus, according to Atomism, the differences between *A* and *K* are to be attributed to different distributions of the same atoms within an unchanging void. Classical Mechanics accentuates the parity between *A* and *K* by showing that the process leading from *A* to *K* is reversible. A reversal of all the velocities in *K* would lead back to *A*.

Every reduction in the number of atoms, every theory proposing, say, that the atoms differ only in mass and charge, is hailed by science as a great step forward. This is because each such step involves a reduction of diversity, *i.e.*, an increase in homogeneity. In the same spirit Einstein proposed to define simplicity in terms of the paucity of the primitive elements occurring in a given theory. In Meyersonian language, the number of primitives is a measure of diversity. Already at this stage, we recognise the close affinity which exists between Meyerson's principle of identity on the one hand and, on the other, Einstein's unity principle which provided the heuristic of Unified Field Theories.

The progress of science is seen by Meyerson as a struggle between the mind which tries to impose identity and a differentiated reality which resists such an imposition. The mind cannot completely succeed in denying all diversity; but some aspects of reality allow of subsumption under the principle of identity; *which* particular aspects can be determined only *a posteriori*, *e.g.*, by trial and error methods. The general identity principle however is *a priori*.

(*b*) *Legal Explanations*

The identity principle applies at two distinct levels: at the second-order level of laws and at the first-order level of things. This gives rise to what Meyerson calls legal explanations and causal explanations respectively. Legality means that the *same* law applies irrespective of place and time, *i.e.*, the form of the law is both a-spatial and a-temporal. There is thus permanence or constancy of form: many diverse phenomena illustrate, or manifest, the same underlying law which constitutes their Platonic Form. For example, the falling stone, the moon revolving around the earth and the passenger thrown forward in a decelerating train are all instantiations of one and the same gravitational law. The permanence of the rule of law enables us to predict, hence has high survival value. According to Meyerson, every animal species needs for its survival the ability to anticipate facts. Thus he gives an evolutionary reason for the *a priori* character of the principle of identity as applied to laws: this principle is *a priori* in the sense of being inborn in a species which needs it for survival.

(*c*) *Causal Explanations and Conservation Laws*

The human mind goes much further than asserting just the permanence of natural laws; it searches for causal explanations, which Meyerson defines as those under which certain substances, or things, are conserved; *i.e.*, the total quantity of these substances remains unchanged through time. Conservation laws thus flow from the principle of identity *qua* negation of the diversity of things in time. I have said that Meyerson calls this version of the principle the causal one: the effect is equal to, *i.e.*, it contains as much substance as, its cause.

Meyerson describes the way in which we construct conservation laws as follows. The mind fixes on certain processes in nature which it hypostasises; *i.e.*, it turns these *processes* into *substances* or *things* whose total quantity it then assumes to be constant. For example, commonsense anchors sense-data in material objects which are supposed to perdure in time. This commonsense hypothesis forms the basis of the classical law of the conservation of mass. Another example is that of the principle of inertia. For Aristotle, motion was change and velocity a ratio describing this process of change. Multiplying this ratio by mass, Descartes turned it into a *thing*, which he called quantity of motion and whose sum total in the world remains constant. This constancy was derived from God's immutability, which is another name for the principle of identity. Similarly, energy is hypostasised motion-cum-position within a field of force (kinetic and potential energies); this automatically leads to a new conservation law, the conservation of energy law. Note that in classical physics matter is the primary substance, while momentum and energy derive their existence from matter (and charge) in motion. Special Relativity puts matter and energy on a par; so we can suppose that there exists a unique substance which our senses apprehend sometimes as radiant energy, say, and sometimes as hard impenetrable matter. Special Relativity thus enhances the status of energy by making it a constituent part of physical reality. This establishes a link between the Special Theory on the one hand and, on the other, General Relativity where energy considerations played an important part in the discovery of the correct field equations. In connection with Relativity however, we should sound a note of warning: substances are generally described not by single quantities but by tensors; the fact that tensor equations are frame-independent reflects the absolute character of substances; conservation is expressed by the vanishing of the tensor's divergence.

(*d*) *Causality, Geometry and Panmathematicism*

Let us go back to the problem of determining what Meyerson means by 'rational explanation'. In view of the above distinction between legality and causality, one might be tempted to equate 'rational' explanations with 'causal' ones, *i.e.*, with explanations which maximise conservation laws. That this is not the case can be seen from Chapter 8, where Meyerson maintains that General Relativity applies to, but does not account for, electromagnetic phenomena:

> Thus Einsteinian Relativity applies without exception to everything pertaining to the action of electricity; however, this does not mean that it explains these phenomena in the sense that it explains gravitation. For in the case of gravitation, as we have just seen, once one has established the presence of a body—that is, the existence of a particular structure in that part of space—all gravitational phenomena logically follow from it. As soon as electrical phenomena intervene, on the other hand, we must assume a particular activity in Einsteinian space that does not follow from the nature of that space alone. (pp. 81–2)

Let us first note that classical electromagnetism obeys strict conservation laws expressed by the vanishing of an ordinary divergence. Secondly, Maxwell's equations can be written in a generally covariant form which exhibits the effect of gravity—*i.e.*, of curved space-time—on electromagnetic processes. Conversely, electromagnetic energy enters into the tensor $T_{\mu\nu}$ which occurs in Einstein's equations.

$$R_{\mu\nu}-\tfrac{1}{2}g_{\mu\nu}R = -kT_{\mu\nu}.$$

In other words, electromagnetic energy modifies the curvature of the manifold. This is why Meyerson had to admit that 'Relativity applies without exception to everything pertaining to the action of electricity'. Thirdly and most importantly, both Einstein and Eddington realised how difficult it was to define gravitational energy. Given a nonzero curvature, there exists no energy tensor whose ordinary divergence vanishes everywhere. One can define a matrix which possesses this property and whose discovery actually led Einstein to the correct field equations. But this matrix is frame-dependent, so there arises the question whether it can denote anything objective. Meyerson was certainly aware of these problems, which were discussed at length both in Einstein's 1915 articles and in Eddington's 'Mathematical Theory of Relativity' [1923], a work often mentioned in the 'Relativistic Deduction'. All this goes to show that, in Meyerson's opinion, the advantage of General Relativity as a framework for the whole of physics lay, not in its superior conservation laws, but in its exclusively geometrical treatment of gravity. This is confirmed by the above quotation.

Meyerson rightly claims that Weyl and Eddington, through trying to construct one geometry from which all aspects of physical reality would flow, were both continuing Einstein's work and striving for greater rationalisation. We cannot therefore equate 'rational' with 'causal', although a good theory should, according to Meyerson, contain conversation laws in one form or another. In line with the last quotation can we perhaps identify 'rational' with 'geometrical'? Again, this cannot be the case, despite the fact that Meyerson considers geometrisation as a major step towards the 'rationalisation' of Nature. At various points of the 'Relativistic Deduction' (*cf.* pp. 31, 116 and especially 151) Meyerson maintains that, because geometry possesses certain qualitative aspects, it cannot be entirely rational; that precisely by virtue of this qualitative character, geometry is both irreducible to analysis and better suited to capturing reality than are algebra and arithmetic. It gradually emerges that, for Meyerson, rationality consists in what he himself called 'panmathematicism': only the strictly mathematical—and this excludes geometry—is rational. In order to explain how Meyerson arrived at this seemingly arbitrary conclusion we have to go back to the Hegelian roots of his thought. By examining the long sections devoted in 'De l'Explication dans les Sciences' to Hegel's and Schelling's philosophies, I shall try to show that Meyerson was an *inverted* Hegelian who shared the Master's views about both formal logic and mathematics (*cf.* Meyerson [1921], Chapters XI–XIV).

(*e*) *Hegel's Heritage*

As is well-known, Hegel tried to derive the essential features of reality from *a priori* principles. He found that formal logic was empty and that mathematics was despicably weak, hence unable to account for the qualitative aspects of Nature. He consequently proposed to go beyond abstract thought towards a dialectical logic rich enough to explain reality.

While profoundly admiring Hegel, Meyerson had to admit that Hegel's attempt had misfired, at least as far as the physical sciences were concerned. Today, we would say that Hegel's logic is overfruitful in that it yields every proposition together with its negation. Yet, in 'De l'Explication dans les Sciences', Meyerson tried to rescue some components of Hegelian 'logic'. He maintained that, in order to yield knowledge, the statement $A = A$ had to be non-tautological in the sense that the A on the right hand side of the equation was not identical with the A on the left side. He argued that the two A's were the same from one viewpoint but different from another. More convincingly, he claimed that the equality $a^2-b^2 = (a-b)(a+b)$ is not an identity. Although both sides of the equation yield the same result when numbers are substituted for a and b, the operations performed on the two sides remain different (*cf.* Meyerson [1921], Chapter V, pp. 131–6). Clearly, Meyerson was making a distinction similar to the Fregean one between reference and sense. But this approach did not really help him in his attempt to assert the equality of cause and effect; for here the cause and the effect differ as regards both sense and reference (*cf.* Meyerson [1921], Chapter V, p. 149).

Meyerson gradually took cognizance of the fact that Hegel had completely failed to exert any influence on the progress of science. Whereas a modern scientist might admire Descartes despite the latter's many false hypotheses, the same scientist could not but view the Hegelian system with bewildered disbelief (*cf.* Meyerson [1921], Chapter XI, pp. 22–3). Even Schelling had more success than Hegel in that Schelling's Naturphilosophie, being more *a posteriori* and hence more in line with scientific practice, fruitfully influenced a first-rate physicist like Oersted (*cf.* Meyerson [1921], Chapter XI, p. 27).

(*f*) *Meyerson's Irrationalist Solution*

Meyerson accepted the Hegelian thesis that science remains rational only as long as it reconstructs nature *a priori*, *i.e.*, independently of any external input. He expressed himself strongly on this point. Following Goblot, he required science to explain 'logically' why things are what they are and not otherwise (*cf.* Meyerson [1921], Chapter III, p. 83). He also agreed with Hegel that mathematics is near-empty. Yet, the sciences seemed to develop by systematically applying the two disciplines judged by Hegel to be contemptibly weak, namely ordinary logic and ordinary mathematics. So there arose the formidable problem of explaining how science progresses by using analytic principles. (I shall henceforth call a proposition analytic if it follows from logico-mathematical axioms, *e.g.*, from the set-theoretic axioms. Needless to say, such axioms are—at least physically—uninterpreted.)

Meyerson's solution was that rational explanation, wherever it operates, 'destroys' reality. In other, less dramatic, words: analytic propositions do not grasp reality at all. But we know that science is realistic. Reality must therefore reenter the scientific picture through channels which are, by definition, irrational (or non-rational). Some of these irrational elements have already been listed: sensation and observation, physical interpretation, empirically ascertained regularities, *etc.* (see (I), (1)–(5) above). Meyerson's position is strikingly similar to that of Eddington for whom relativity theory is largely *a priori* and, in as far as it is *a priori*, does not refer to the physical world at all (*cf.* Eddington [1920], pp. 196–7 and [1923], p. 120). It will be shown that these views are diametrically opposed to Einstein's.

(*g*) *Meyerson's Conception of the Apriori*

In Chapter 23 of the 'Relativistic Deduction' Meyerson investigates the possibility of an evolution of human reason. He interestingly concludes that *the demarcation between the a priori* and the *aposteriori* had shifted over the years, but that the shift had always taken place in the same direction. Certain propositions, like the Euclidicity and the three-dimensionality of space, which had once been regarded as sacrosanct *a priori* principles, were gradually shed, *i.e.*, they were surrendered to the empirical domain; they were made to stand or fall with the experimental confirmation or refutation of the hypotheses containing them. Of course, this did not prevent physicists from later trying to reintegrate these propositions into the *a priori* sphere: Hermann Weyl tried to derive the four-dimensionality of space-time from *a priori* principles but, as Eddington showed, his attempt failed (*cf.* Eddington [1923], p. 213). Thus the *a priori* component of thought is gradually impoverished, *i.e.*, it becomes more analytic (where, as usual, strict analyticity means truth under an arbitrary interpretation of the descriptive terms). Meyerson in effect admits varying degrees of rationality and hence of analyticity. For example, we have seen that geometry is taken to be less rational than arithmetic or algebra. This can be understood to mean that the distance formula:

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu,$$

where the $g_{\mu\nu}$'s are given functions of the coordinates, does not logically follow from the mathematical axioms (*e.g.*, from analysis or set theory).

Meyerson's approach can be pushed further to yield the modern view that certain mathematical postulates, like the Axiom of Choice and the Continuum Hypothesis, should be considered neither analytic nor *a priori.* Depending on whether one accepts these postulates or asserts one of their contraries, one obtains one or other from among many possible set theories. One could go even further. From a Meyersonian point of view, there seems to be no reason why certain Boolean laws of ordinary logic should not be jettisoned if the physical theories which make use of them fail to provide a coherent account of the facts. Meyerson is thus very modern in that quantum logic would not—or rather, should not—have shocked him.

(*h*) *The Defects of Meyerson's Position and its Reinterpretation*

Clearly, Meyerson's philosophy of science, as it stands, has many shortcomings. I shall now try to expose these shortcomings, then offer a reconstruction which remedies the most obvious defects of Meyerson's position while preserving its essential core.

Meyerson's Hegelian treatment of equality, whereby $A = K$ (from one point of view) and $A \neq K$ (from another), is clearly untenable. Fortunately, in order to formulate his version of the causality principle, he does not have to assert, as he nonetheless does, an equality between the antecedent A and the consequent K. A and K need be neither equal nor, if interpreted as propositions, logically equivalent; they only have to be equivalent in the mathematical sense of an equivalence relation. Going back to the hypostasisation procedure described on p. 97, we can say that the mind singles out some process common to A and K; it then transforms this process into a substance, which will be denoted by $\phi(A)$ and $\phi(K)$ in the cases of A and K respectively; it finally expresses a conservation law through the equality: $\phi(A) = \phi(K)$. This equation, which can also be written as $A \sim K$, obviously expresses an equivalence relation between A and K. $\phi(A)$ may for example represent the total mass, total momentum or total energy in A. Thus $\phi(A) = \phi(K)$ asserts, not that A equals K, but simply that A and K are equivalent as regards mass, momentum or energy.

Much more serious is the difficulty which Meyerson finds with the physical interpretation of mathematical entities. In this problem area his whole approach proves wrong-headed. Misled by Hegel, he claims in effect that scientific speculations start out by being abstract mathematical hypotheses which are then given a physical interpretation. (According to Meyerson, the 'relativist' may for a long time remain oblivious to the fact that his symbols will ultimately have to denote something physical.) Only the formal propositions are rational. As for the subsequent interpretation, it constitutes an irrational moment in the process of theory-construction.

We can now answer in the affirmative the question, posed on p. 95, whether items (1)–(5) have anything in common: they are all designed to take us away from a supposedly abstract mathematical system towards a realist physical theory. For example: (1) and (3) connect the system with observation and with empirical laws respectively; (2) provides for the overall physical interpretation of mathematical symbols and of theoretical terms.

Needless to say, this is not how scientists actually proceed. Right from the beginning, their speculations, though not necessarily arising from observed facts, are interpreted statements about physical reality. This is why, despite his great admiration for Meyerson, Einstein wrote in his review of the 'Relativistic Deduction':

> I cannot namely admit that the assertion that the theory of Relativity traces physics back to geometry has a clear meaning. One can with better justification say that, with the theory of Relativity, (metrical) geometry has lost its special existence vis-a-vis the instances of law-likeness which have always been denoted as physics . . . Even before the proposal of the theory of relativity it was unjustified to consider geometry vis-a-vis physics as an 'a priori' doctrine. (cf. p. 254)

Let us recall that, in 'Geometry and Experience', Einstein made a distinction between two kinds of geometry: the mathematical variety and the physical one. Any mathematical geometry, being a formal system, has to satisfy a mere consistency condition. As for physically interpreted geometry, it constitutes one of the oldest empirical theories and is therefore subject to experimental strictures (*cf.* Einstein [1934], pp. 119–27). Speaking about Relativity, one can maintain with equal legitimacy that it physicalises geometry and that it geometrises nature. The reason for which Einstein protested against Meyerson's concept of geometrisation was that Meyerson gave this concept its abstract mathematical sense .

Going back to the nature of physical theories, it is clear that they are *interpreted* synthetic propositions. In fact, the more phenomena a theory sets out to explain, the more powerful and hence the more contingent it becomes; for analyticity is trivially hereditary in that it is passed on by every analytic statement to each of its logical consequences. Yet Einstein often claimed that science was capable of explaining the world *a priori* (*cf. e.g.*, Einstein [1934], pp. 116–17). By virtue of what has just been said, 'a priori' cannot be equated with 'analytic', even though 'analytic' has been extended to mean 'either strictly analytic or mathematical'. Can 'a priori' be given any meaning at all?

It is well known that Einstein required theories to satisfy two criteria: an external criterion of consistency with the facts and a so-called criterion of internal perfection (*cf.* Einstein [1949], pp. 21–3). I do not intend to go into the various meanings which 'internal perfection' has been given, *e.g.*, simplicity, unity, organic compactness, paucity of primitives, *etc.* Suffice it to say that, unlike analyticity, internal perfection is *not* a hereditary property. On the contrary, the more seemingly disconnected its empirical consequences, the more internally perfect a hypothesis is taken to be. Let us define a good *a priori* theory to be an internally perfect universal proposition which explains the facts in a non *ad hoc* way. In particular, such a theory should not have been doctored to yield known experimental results; it should rather have arisen from general speculation about the nature of the physical world. This definition is admittedly vague, but no more vague than Meyerson's concept of rationality. Moreover, had Meyerson adopted this definition of 'a priori' *and* equated 'a priori' with 'rational' (he anyway did the latter), then his position would have been immensely strengthened. He would not have needed to resort to irrationality in order to inject life into supposedly empty rational explanations. Physical meaning and experimental relevance, instead of having to be imported into a theory, would have been built into it right from the start.

Meyerson was a fallibilist hypothetico-deductivist who subscribed to what he called the negative dogmatism of science: science is certain only about the hypotheses it rejects, not about those it temporarily accepts (pp. 225–32). The proposed redefinition of 'a priori' would make sense of this falsificationist methodology: 'rational' explanations, instead of being abstract, hence untestable, would now become contingent and empirical, albeit internally perfect; their consequences can thus be pitted against experience, hence possibly refuted. This would provide a rationale for Meyerson's negative dogmatism.

An example of a good *a priori* hypothesis is Weyl's unified field theory. Such a theory, being an all-embracing physical geometry, provides a maximal illustration of the Identity Principle. Since geometry is now looked upon as an interpreted system, it must be *about* something; this something can be regarded as one entity, a kind of universal ether or field. The different sensations, which reveal the presence of this field, are separated from one another by unbridgeable gaps in a way that the continuously varying intensities of the same field at different points are not so split. Unified field theories thus illustrate the Identity Principle by substituting a single medium for our heterogenous perceptual domain. Meyerson does not seem to notice that this argument, which he himself put forward (in the opening chapter of the 'Relativistic Deduction'), is predicated on the assumption that the geometry he has in mind must be a physically interpreted one.

One irrational jump however remains, namely the transition from the kinematical processes postulated by Relativity to the phenomenon of sensation as such. But this so-called mind-body problem is not peculiar to Meyerson's philosophy. Meyerson may well be right in affirming that this gap will never be bridged in an intelligible way (*cf.* Meyerson [1921], Chapter VI, pp. 184–5).

So far we have spoken only of Items (1)–(3), pp. 94–95. Let us now turn to (5) which, in my opinion, arises from a misconception of the nature of mathematics. Throughout most of Meyerson's scientific life, the mathematical methods used in physics were limited to the analysis of continuous and differentiable functions. It looked as if respectable mathematics accommodated only well-behaved continuous entities. Thus, if one identifies the 'rational and a priori' with the 'mathematical', then the quantum jumps postulated by atomic physics will automatically fall outside the domain of rationality. This is why Meyerson regarded quantum theory as a very serious obstacle to the 'rationalisation' of Nature and why, in his 'Réel et Déterminisme dans la Physique Quantique' [1933], his whole position threatened to collapse. Quantum Theory looked like the ultimate revenge of Nature against the Identity Principle which the mind tries to impose on the phenomena.

Meyerson need not have worried that much. Not only are most real-valued functions non-continuous but, after Dedekind, he ought to have known that the continuum itself is built from the discrete set of fractions. (Non-relativistic) Quantum Mechanics has not only been elegantly mathematised, it has also been shown to obey strict conservation laws. It has moreover been given acceptable realist interpretations. Quantum Physics is thus far less 'irrational' than Meyerson feared. Of course, there remain the problems of determinism and of sharp spatio-temporal location. In the 'Relativistic Deduction' Meyerson does not discuss the determinism problem at any length, probably because he looked upon determinism as a fundamental postulate. There does not however appear to be any direct link between determinism and rationality. But precise spatial position at a sharply defined moment is certainly taken for granted whenever Meyerson speaks of the conservation of substance. It seemed obvious to him that, in order to assert meaningfully that a thing perdures through time, its boundaries at a precise instant should be fully determinate. Only in this—limited—sense did Quantum Theory prove refractory to the Identity Principle.

We shall finally deal with item (4), which turns out to be so important that the following section is wholly devoted to it.

3 THE IRREVERSIBILITY OF TIME

The following two theses are conflated both by Meyerson and by Milič Capek in his otherwise excellent introduction to the 'Relativistic Deduction'.

(α) Mathematical physics treats—or rather: would like to treat—time on a par with space; it tends to *spatialise* time. More precisely: time is nothing but an extra dimension denoted by an abstract quantity t which is in every way similar to the spatial coordinates x, y, z.

(β) Time has no preferred direction; so the relations 'before' and 'after' possess no objective status.

According to Meyerson, the ascertained unidirectionality of time, being tantamount to a denial of the possibility of complete spatialisation, is one further obstacle against pan-mathematicism, *i.e.*, against the rationality of Nature. (pp. 257–8)

Let us assume, for the sake of argument, that the implication $(\alpha) \Rightarrow (\beta)$ holds. It can in fact be argued for as follows: since the choice of a positive direction on a geometrical axis is a matter of convention and since the temporal dimension is supposed to be in all respects similar to a spatial coordinate, the alleged unidirectionlity of time possesses no objective status. However, the converse $(\beta) \Rightarrow (\alpha)$ does not hold. Relativity in effect falsifies $(\beta) \Rightarrow (\alpha)$: both Einstein and Meyerson correctly pointed out that, since the signature of $(g_{\mu\nu})$ is $(1, -1, -1, -1)$, Relativity does not treat time and space on a par. Weyl expressed this state of affairs by saying that the spatio-temporal manifold is $(3+1)$-dimensional rather than 4-dimensional. Yet, in Special Relativity, the choice of a new frame of reference may reverse the time-order of two (space-like) events. The existence of particles moving with superluminal velocity, a possibility not excluded by Relativity, makes the situation even worse. Thus time-ordering, being frame-dependent, has no objective significance in the special theory; and Gödel showed that this conclusion can be extended to the case of General Relativity (*cf.* Gödel [1949]).

It is true that, even before Relativity, all fundamental classical theories were time-reversible. However, the time interval Δt was, or could be made into, an invariant. Of two events E_1 and E_2, E_1 could be said to precede E_2 objectively if $\Delta t = t_2 - t_1 > 0$; hence the *possibility* of defining an objective time-ordering over and above all classical time-reversible hypotheses. Gödel showed that Relativity destroys this possibility and, as Capek himself admits, Einstein more often than not, and right to the end of his life, agreed with Gödel. One is tempted to say that, in Relativity, events are given all at once; except that this 'at once' should be understood in a purely figurative sense. It does not refer to a supertime in which events are placed. On the contrary: time is written *into* each event, or rather: every frame splits the events constituting the universe along one temporal and three spatial axes. But the universe as such is not created in time, it simply *is*. It follows that, although space and time are not treated exactly on a par, one aspect of the dissimilarity between them, namely the unidirectionality of time, has been eliminated. This represents an increase in homogeneity as required by the Identity Principle. Putting it a little differently: the block-universe interpretation, which subsumes Becoming under Being and is in effect a—partial—return to the Parmenidean Sphere, ought to have pleased Meyerson since it strongly supports his central thesis. It shows that Relativity, as a development from Classical Physics, is a movement towards greater 'identity'. But, as already mentioned, Meyerson was misled by Hegel into believing that he, Meyerson, subscribed to a formal rationality which would destroy reality. So Meyerson went on wrongly insisting that time-irreversibility should, and did, set definite limits to the success of mathematisation and hence of rationality.

CONCLUSION

The 'Relativistic Deduction' is a classic which ought to be read in conjunction with, and as a corrective to, another great classic on the foundations of Relativity: Reichenbach's *Philosophy of Space and Time*. Unlike most members of the Vienna Circle, Meyerson knew that science needed both a realist metaphysic and a non-inductivist methodology. It is a pity that, also unlike the Vienna Circle, he did not follow the developments of contemporary logic but chose instead to grapple with Hegel's sterile idealist system. This choice, whether deliberate or instinctive, may be due to the fact that Meyerson spent his youth in the 19th century; it anyway explains why it was necessary to carry out a 'rational' reconstruction of his position, which in no way diminishes his achievement.

ELIE ZAHAR
Department of Philosophy
The London School of Economics

REFERENCES

EDDINGTON, A. S. [1920]: *Space, Time and Gravitation*. Cambridge University Press.
EDDINGTON, A. S. [1923]: *The Mathematical Theory of Relativity*. Cambridge University Press.
EINSTEIN, A. [1934]: *Mein Weltbild*. Ullstein, Frankfurt.
EINSTEIN, A. [1949]: 'Autobiographical Notes' in P. A. Schilpp (*ed.*) *Albert Einstein: Philosopher-Scientist*, pp. 1–95. Tudor Publishing Company, New York.
GÖDEL, K. [1949]: 'A remark about the relationship between Relativity Theory and Idealistic Philosophy', in P. A. Schilpp (*ed.*) *Albert Einstein: Philosopher-Scientist*, pp. 557–62. Tudor Publishing Company, New York.
MEYERSON, E. [1921]: *De l'Explication dans les Sciences*. Payot, Paris.
MEYERSON, E. [1933]: *Réel et Déterminisme dans la Physique Quantique*. Hermann, Paris.
REICHENBACH, H. [1958]; *The Philosophy of Space and Time*. Dover Publications, New York.
SCHILPP, P. A. (*ed.*): *Albert Einstein: Philosopher-Scientist*. Tudor Publishing Company, New York.

PHILOSOPHY, PSYCHOLOGY AND PSYCHOANALYSIS*

In *The Philosophical Foundations of Psychoanalysis* Grünbaum [1984] has launched a powerful attack upon the epistemology of the notion of repression as pathogenic for neurosis in psychoanalytic theory. He shows that, inter alia, even if the clinical data themselves which were used by Freud are accepted as a reasonable basis for a scientific theory (an assumption incidentally with which many psychologists would strongly disagree), even then they cannot stand as support for the theory. Without this support for repression, the cornerstone of the theory, the whole edifice of psychoanalysis collapses into ruin, the argument runs, and the question arises as why as philosophers or psychologists we should rationally concern ourselves with psychoanalysis at all.

---

* Review of Adolph Grünbaum [1984]: *The Foundations of Psychoanalysis: a philosophical critique*. University of California Press. *xiv*+309 pp. (Hardback $16.95) (ISBN 0–520–05016–9).

In delivering this philosophical coup de grâce, Grünbaum also attacks with no little success the modern attempts to rescue psychoanalytic theory by considering it not as science but as hermeneutics. These too he shows to be incoherent, riddled with false assumptions and poor logic inferior as examples of philosophy to Freud's own attempts to make his theorising watertight. Furthermore, although by no means supporting Freudian theory, as should by now be obvious, Grünbaum exposes problems and difficulties in Popper's somewhat simplistic approach to the rejection of Freudian theory on the basis of its unfalsifiability.

In brief Grünbaum has exposed severe, indeed, destructive weaknesses in the epistemology of repression by arguments which I fully accept as sound. Nevertheless Grünbaum's devastating conclusions concerning Freudian theory I think by no means necessarily follow from his arguments. To make this point it is clearly necessary to delineate the nub of Grünbaum's logic. When this has been done, I shall put an alternative viewpoint which somewhat surprisingly Grünbaum fails to discuss. In addition I shall point out flaws in Grünbaum's work which stem from his philosophical background and his inevitable limitations, which should be recognised by readers, in psychological research, both in the design and evaluation of experimental work.

The quintessence of Grünbaum's case is that contrary to Popper's claims the clinical psychoanalytic method and the causal inferences based upon it are flawed epistemically but for reasons other than non-falsifiability. Freud (*e.g.* [1937]) had argued that the consilience of clinical inductions supported his theorising and this Grünbaum demonstrates just will not do. In other words, Freudian theory is falsifiable but not by the psychoanalytic method. Instead extra-clinical, experimental verification is required.

It is interesting and instructive to note Grünbaum's rejection of the Popperian claim, because, as I shall show, some of these arguments, and they are powerful, may be turned against Grünbaum's later conclusions.

As Grünbaum points out, Popper's claim that psychoanalytic theory cannot be falsified rests upon what many might consider to be an insecure basis, namely that philosophers cannot think of ways to test them. More importantly there is the logical issue as to the proof that Popper offers for his statements, that Freudian theory is devoid of empirically testable consequences. How did Popper convince himself, Grünbaum asks (p. 113) that the infinite consequence class of Freud's theory contains no testable members at all. In fact Popper attempts no such proof, an exceedingly difficult demonstration which might indeed be impossible, but rather relies on enumerative induction, a procedure which is insufficient for the task.

What is noteworthy about this rejection of Popper's case is the first argument: Freudian theory cannot be falsified because philosophers cannot immediately see how it might be done. As I shall discuss in more detail later in this paper, this is exactly analogous to an important part of Grünbaum's arguments. He rejects the extra-clinical, experimental evidence cited by some writers as supporting Freudian theory because he, a philosopher, does not believe it does so. This is to be hoisted on the Popperian Petard.

There is a further neat argument against Popper's view of Freud which Grünbaum uses. Popper claims Freudian theory excludes no behaviour, hence can explain anything. However, if the theory excludes nothing, it can explain nothing for, as was argued by Spinoza to explain deductively is to exclude. To assert *P* is to deny any proposition with which it is incompatible.

Actually this rejection of the Popperian argument is very much by the way. However, it is important to clarify the grounds for rejecting what has, in the field of psychology, at least, become received wisdom, since it is also an important aspect of the argument to be used in the rebuttal of Grünbaum's thesis.

Grünbaum, who is nothing if not careful with words, has entitled this book *The Foundations of Psychoanalysis*, because he argues that his demonstration of the epistemic weakness of the clinical method for demonstrating the link between repression and neurosis undermines psychoanalytic theory as a whole. Freud was aware of the philosophical problems in his method (as many of his apostles and critics seem not to be) but despite his efforts to counter these, he failed.

First I shall examine Grünbaum's argument demonstrating this failure. At the outset I may say that it is a convincing case, and one that many apologists for Freud have not yet answered. Freud, according to Grünbaum, made five claims concerning the probative value of clinical data, culled from the psychoanalytic couch, as a basis for psychoanalytic theory, and for the efficacy of psychoanalytic therapy. These claims answered criticisms that would, prima facie, arise in any detailed consideration of psychoanalysis by anyone other than committed psychoanalysts. These claims which, if true, would render to psychoanalytic therapy and theory the importance which its adherents believe it due, are: that clinical data are not contaminated by suggestion; that other therapies work through suggestion; that it is possible to validate causal claims by retrospective methods using psychoanalytic data; that it is possible to attribute success in therapy to psychoanalysis without the use of control groups and finally that once repression and concomitant distortions are removed, introspections become reliable.

These, as must be obvious, are no minor claims and Freud used what Grünbaum has called the tally argument to support them. However, this tally argument fails to withstand Grünbaum's exacting scrutiny, hence these claims are undermined and if these are denied with them goes the cornerstone of psychoanalysis, repression as pathogenic for neurosis; for on this depends the third claim that causal links can be uncovered in retrospective studies.

Since, essentially, the book turns on the destruction of the tally argument, I shall now examine this and the objections to it which have been raised by Grünbaum.

In the psychoanalytic conception of the dynamics of successful psychotherapy there is a fundamental premise: conflicts are only solved if the interpretations tally with what is real in the Patient (Freud [1917]). Grünbaum explicates the assumptions in this argument, presented in Freud's paper on analytic therapy. In fact for successful psychotherapy (strictly only of the psychoneuroses) there are two causally necessary conditions:

(1) Only the psychoanalytic method can yield correct insight into the pathogenis of neurosis.
(2) The patient's correct insight is causally necessary for the conquest of neurosis.

Grünbaum refers to this conjunction of claims as the Necessary Conditions Thesis which for simplicity I shall abbreviate in this paper to *NCT*. With this *NCT*, as Grünbaum points out, Freud can use clinical data as support for his theory.

Thus from patients who have been successfully treated by psychoanalysis two conclusions follow, namely that psychoanalytic interpretations of the hidden, unconscious, causes of behaviour are correct and these tally with what is real in the patient and second only psychoanalytic treatment could have wrought this conquest. This is the tally argument, as labelled by Grünbaum, which Freud used to justify the scientific probity of psychoanalytic theory. They are sophisticated points because they answer many of the immediate objections to Freudian theory. However, this tally argument, in the view of Grünbaum, will not stand scrutiny.

It is clear that *NCT* does allow clinical data to be used as validation of Freudian Theory. On the other hand equally clearly *NCT* collapses and thus renders the theory suspect, if psychoanalytic therapy does not work, if other therapies work and if there is no symptom substitution. Much depends, therefore, on the empirical evidence.

Grünbaum, then, goes on to cite modern studies into the effectiveness of psychoanalytic therapy, spontaneous remission and the psychotherapeutic efficiency of other therapies, arguing that their findings throw considerable doubt upon (even if they are not entirely contrary to) the basis premise of *NCT*. This in turn cast doubt on the clinical evidence Freud used and hence the theoretical conclusions themselves.

At this juncture of the argument my first real point of contention arises. So far all has turned, quite properly, on philosophical issues, matters of logic and debate. Grünbaum has written with great skill and a formidable mastery of detail, as befits a renowned philosopher. However, while it is philosophical to claim that empirical, psychological studies are required to support or refute a premise, it is *not* philosophical but *psychological* (in the sense of necessitating psychological expertise) to cite studies and argue that the case is or is not supported. Just as, in the instance of Popper declaring psychoanalysis to be untestable, Grünbaum rightly in my view, questioned his authority, qua philosopher, to make this judgement, so I use the same argument against Grünbaum in declaring the empirical evidence to refute the *NCT*. I am not here disputing the truth or falsity of the claim, simply arguing that Grünbaum has no special qualifications to render his assertions credible. He needs to justify these assertions by reference to the research design and scientific adequacy of the studies. Matters, indeed, of psychological expertise. This is particularly pertinent with respect to the evidence for psychotherapy, which among psychologists has long been a matter of contention, see Kline [1981]. For example Grünbaum cites with approval the work of Smith *et al.* [1981] although, to indicate there is little agreement, I regard this methodology as hopelessly flawed (Kline [1983]).

In brief, I am arguing not that Grünbaum's views are wrong (although incidentally I think them so), but that these are contentious matters requiring an expertise which I am not convinced, from background, training and previous publications, Grünbaum possesses and in which I would have to feel confident if I were to accept his thesis, given the lack of detailed argument in support. I am sure that Grünbaum does not, unlike Jowett, feel that what he does not know is not knowledge.

If Grünbaum's citations of recent empirical work which calls into question the efficacy of psychoanalytic therapy, demonstrates that other therapies can be effective and that symptom substitution is by no means inevitable are taken at their face-value, then the force of *NCT* and the tally argument is blunted: what psychotherapeutic success there is, with psychotherapy, could be done to suggestibility (Freud's original reason for developing the tally argument). However, although this is logically so, logic must be distinguished from actuality. It is possible to run a session of psychoanalytic therapy, quite free from suggestion, even though there is no way of demonstrating this fact. So that even if Grünbaum's arguments undercut the tally argument and *NCT* and thus render the theory dubious and the data doubtful, it is still viable to argue that the theory is tenable, although owing to the problem of suggestability there can be no definite confirmation. It does not follow from Grünbaum's arguments that psychoanalytic theory should be abandoned. Grünbaum's conclusion from the tally argument is that Freud's attempt to justify clinical data as a basis for psychoanalytic theory fails because the necessary premise of psychotherapeutic success fails. The destruction of the tally argument undermines the data base of the theory.

This is all bad enough from the viewpoint of a convinced Freudian, but in the later sections of *The Foundations of Psychoanalysis*, Grünbaum goes on to show that even if clinical data could be taken at face value as being uncontaminated epistemically, the reasoning by which Freud sought to justify the foundations of his theory, the clinical arguments for repression as pathogenic for psychoneurosis, and its role in parapraxes and dreaming, are themselves so flawed as to render them worthless.

In this study of Grünbaum's work I shall not scrutinise all these separate chapters in detail. The method and approach is similar in all and the counter arguments which Grünbaum never attempts to rebuff are similar for each. Rather I shall concentrate on Grünbaum's attack on Freud's arguments concerning repression in the aetiology of the psychoneuroses, an aspect of psychoanalytic theory described as its cornerstone.

In 1893 Freud and Breuer argued that the undoing of repressions led to therapeutic success. This was the basis for the central dynamic significance of unconscious ideation in psychoanalytic theory. Without this attribution, it would have been impossible to postulate the aetiological role of repression in psychoneurosis.

However, by 1896 failures with patients forced Freud to repudiate the decisive therapeutic efficacy of removing repression of the occasioning trauma. As he recalled in his Autobiographical Study, Freud [1925] came to realise that the personal emotional relationship between patient and therapist was the stronger influence. This repudiation of the role of lifting repressions in symptom removal in fact, according to Grünbaum, negated the very reason that Freud and Breuer had invoked for claiming that repression was pathogenic at all for psychoneurosis. Thus after 1896, psychoanalytic theory was essentially incoherent.

Why, then, did Freud not simply abandon the theory of the Pathogenicity of Repression? This is the next question raised by Grünbaum and one incidentally which he never answers, except perhaps by implication, namely that Freud did not fully appreciate the extent to which therapeutic success supported the aetiological role of repression in psychoneurosis.

However, there is a further disastrous sequela of this admission that lifting repression is not sufficient to remove symptoms. For Grünbaum argues that the demise of the therapeutic justification for the aetiological role of repression is in effect the coup de grâce for free association as a method of aetiological enquiry. If psychotherapy is not successful the value of free association as an epistemic avenue to pathogens and hence as research tool is rendered nugatory. As Grünbaum concludes the attribution of therapeutic success to the removal of repressions remains to the day the 'sole epistemic underwriter of the purported ability of the patient's free associations to certify causes'.

In such a manner Grünbaum effectively demonstrates that in the terms of Freud's own arguments there is little support for the theory of repression as pathogenic in psychoneurosis or for its purported role in parapraxes and dreaming. He has destroyed the clinical evidence by revealing its incoherence and its reliance on psychotherapeutic success which even Freud himself did not claim. However, he allows for the fact that extra-clinical evidence (namely experimental) may in the future come to show that Freud, despite the epistemic weaknesses, was not so wide of the mark. He dismisses what experimental evidence there is as failing to support psychoanalysis (p. 278) and the psychoanalytic corpus.

As it stands Grünbaum's analysis of Freudian theory is a devastating critique. The epistemic flaws and incoherence of the clinical theory are exposed for all to see. If there is any truth in Freudian theory it must be serendipitous. This book raises the question (emphatically answered by Eysenck and his colleagues) as to what is the point of any longer continuing to study psychoanalytic theory as a source of scientific knowledge of human behaviour. This is particularly so given that Grünbaum has dismissed the extra-clinical experimental evidence.

Despite the quality of the arguments and the minute, talmudic scrutiny which Grünbaum has employed in his destruction of Freudian theory, I reject his conclusions and I shall devote the final section of this article to making this case. There are other viewpoints to Freudian theory which even though relevant to his thesis Grünbaum has chosen to ignore, or is Oxford so far beyond the diocese of the Cathedral of Higher Learning in Pittsburgh as to be beyond his ken?

What Grünbaum has achieved in this book is a demonstration of the epistemic weakness in Freud's claims of the importance of repression in psychoneurosis, dreams and parapraxes. In this demonstration he casts doubt on the ability of free association to be a useful tool of enquiry. However, although repression is a cornerstone of psychoanalytic theory, his scrutiny is restricted to repression.

This restriction to repression is important because by p. 278 he refers to his work as bearing on psychoanalysis and the psychoanalytic corpus. This is inaccurate. It bears only on repression.

This early transition from a part to the whole is relevant to my first argument against Grünbaum. Farrell (*e.g.* [1961], [1981]) to whom incidentally there is no reference, has often made the point that it is simplistic to talk of psychoanalytic theory. Rather it is possible to conceive of Freudian theory as a set of hypotheses, of which some may be true, others false. The scientific study of Freudian theory, therefore, on this view, entails sifting through these hypotheses and subjecting them to experimental verification. In this way the insights of psychoanalysis which are in accord with the evidence can be retained and the dross rejected—the Popperian approach to the scientific analysis of Freudian theory. An objection was raised to this method by Martin [1964] who argued that to restate the hypotheses in testable form is to change them and hence the problem of relevance arises. While this is sometimes true or could be true, it need not be so. Indeed in many instances it is possible to take the hypothesis direct from Freud's writing. In brief Farrell advocates the casting of Freudian theory into a set of scientifically testable hypotheses so that those resisting falsification can be found.

On this view, Grünbaum has scrutinised (and found wanting) but little of the theory, albeit his work was concerned with an important and central concept in psychoanalysis. On this view, the incoherence and incompatibility of one part with another is accepted. Farrell regards psychoanalysis as a pre-scientific theory. Part of the falsification and testing might lead to a tighter theory with greater consistency and coherence. Now it is to be noted that Farrell's position is purely philosophical. He suggests what experiments might be done. Certainly, as befits the Wilde Reader in Mental Philosophy at Oxford, he ran no experiments. Certainly when he does comment about experiments he makes it clear that his judgements are guided by experimentalists, by psychologists who work in the relevant field. In any case such judgements form a minor part of his work.

If we adopt Farrell's view of Freudian theory, it can be seen that even if Grünbaum's strictures about the nature of the experimental evidence are accepted, all that has been shown is that there is little evidence for the aetiological role of repression in psychoneurosis, parapraxes and dreaming. This is decisively not rejecting the Freudian corpus as a whole.

As it happens a number of experimental psychologists have adopted a viewpoint similar to that of Farrell of which the best known are Sears [1943] or Fisher and Greenberg [1977] or explicitly that of Farrell, as in the case of the present writer (Kline [1972], [1981]). I shall ignore the work of Sears, a small scale pioneer in the validation of Freudian theory and concentrate upon the larger works of Fisher and Greenberg [1977] and myself, together with the reply to the latter by Eysenck and Wilson [1973].

Kline in his two editions of *Fact and Fantasy in Freudian Theory* set out the hypotheses relevant to the main concepts of Freudian theory—oedipus complex, castration complex, superego, for example—and then subjected all objective studies which he could discover to critical analysis, concluding that certain concepts in the theory had received some unequivocal support. In all approximately 1,000 researches were examined. Fisher and Greenberg [1977] carried out essentially the same exercise citing even more references, by including dissertations, but having a far less critical attitude towards the studies, as I have elsewhere shown (Kline [1981]). Despite this there was good agreement over what parts of psychoanalytic theory had received support.

Eysenck and Wilson, on the other hand, took the best 20 studies from the work of Kline [1972], printed them in full together with some brief comments trying to demonstrate that they did not support the theory.

I do not intend to talk about the findings from these large scale surveys, for this is a matter of substantive psychology. Rather I shall make a few points about the approach and methods contrasting it with that propounded by Grünbaum.

In the first place, with objective, quantitative studies of Freudian theory, it is often the case that there is disagreement as to the probative value of the study. Thus Eysenck and Wilson [1973] interpret the same work as I cited [1972] as confirming Freudian theory as failing to do so. This is particularly pertinent since Grünbaum, who is not a psychologist, does cite certain empirical studies. It is, therefore, important to see how such disagreement can occur (assuming for the sake of the argument, that neither of the protagonists is so stupid as to be actually wrong, failing to understand the research, an assumption, perhaps, that is somewhat rash).

One such source of conflict concerns the relevance of the tested hypothesis to Freudian theory. Grünbaum has raised this difficulty which is one that a sound knowledge of psychoanalytic theory should be sufficient to elucidate. A celebrated example, relevant to repression, was the early study investigating whether pleasant words (*e.g.* sugar, sweet) were better remembered than unpleasant words (*e.g.* bitter or sour). There should be little problem in dismissing such a study as irrelevant to the Freudian concept of repression.

Another source of difficulty concerns the psychoanalytic hypotheses but arises not from relevance but from the principle of Occam's razor. For example, in a study of paranoid schizophrenia by Zamansky [1954] which I cited (Kline [1972]) as support for the psychoanalytic theory that latent homosexuality is the cause of the delusions, Zamansky showed that compared with controls the paranoids fixated longer on pictures which were homosexually arousing, such fixation having been shown to indicate desire. However, they did not admit, on questioning, that they preferred homosexual to heterosexual stimuli.

Eysenck and Wilson [1973] argue that an alternative hypothesis is a simpler explanation and hence to be preferred: the paranoids, being suspicious 'of the shrink' would not admit to liking the homosexual pictures. There are two points. First there is no evidence that paranoid schizophrenics are generally suspicious. Rather their paranoia is often encapsulated within a delusional system (Mayer-Gross, Slater and Roth [1967]. In addition a motive such as 'suspicion of the shrink' is entirely post hoc, invoked to explain results that were predicted from psychoanalytic theory. From the viewpoint of a philosopher, the two explanations are probably no different, although why Grünbaum should prefer the latter to the former escapes this writer. To a psychologist of personality who is acquainted with all the myriad theories of personality (of which psychoanalysis is but one) this latter exploration is singularly weak since it is embedded in none of them and is clearly post hoc.

However, there is a further more general, and I think more telling, point about this putative law of parsimony. In some cases Eysenck and Wilson were able to develop some convincing alternative explanations (where to choose between them must be done experimentally rather than by the fiat of however distinguished a philosopher). Over the twenty or so studies which they cited their set of alternative hypotheses might be more simple explanations than psychoanalytic theory and hence to be preferred, excepting the grotesque examples of post hoc explanation. However, if this were to be extended to the far larger number of studies cited by Fisher and Greenberg [1977] or by Kline [1981], this becomes exceedingly dubious.

However, problems over hypotheses are not, in fact, the chief issue. Where there are differences in interpretation, these usually stem from technical matters of psychological research methods. As a philosopher it is perfectly proper to argue that a test of Freud's theory of orality, for example, involves the empirical study of whether oral personality traits cluster together as described in psychoanalysis and whether in some way, if they do so, this syndrome is related to breast feeding, weaning or some other operational definition or oral erotism.

However, how this is to be done and whether it is properly done is a matter for psychologists. It is arrogant of a philosopher to choose one study or another or to judge all equal when, as in this case, there are grievous methodological difficulties which have often not been overcome. How many philosophers know the problems of factor analysis, the artefacts arising from the wrong choice as to the number of factors, from the different methods of rotation, from the variability of subjects, from the failure to reach simple structure, from the lack of hyperplane stuff, from the use of components rather than factors? Yet all these are vital in the assessment of work on oral personality where factor analysis is the logical and most fitting statistical technique.

When I claimed that my book was more critical than Fisher and Greenberg's, it was on account of my concentration on the experimental methods used in studies. I have attempted to discount investigations which are methodologically flawed. Grünbaum, for example, cites the work of Smith and Glass [1970] on the efficacy of therapy. Yet this work which sums together findings from studies regardless of quality is thus necessarily dubious. Similarly it is particularly annoying that Grünbaum should cite the work of Eysenck and Wilson [1973] who set out in full a handful of studies from my first book with approbation when I cannot believe that he has the professional knowledge to make a judgement between the two. For a philosopher to enter this conflict is a mark of prejudice or authoritarianism, deference to the famous.

The alternative approach which I have advocated here, based upon the analysis of the work of Farrell, answers the accusation to my work, raised by Grünbaum, namely that the research which I examined in the study of repression, supported only the fact of repression not its aetiologic significance in neurosis. Of course, this is so, but à la Farrell, this is important. There are several hypotheses: (1) that there is a defence mechanism repression; (2) that repression is of aetiological significance in psychoneurosis; (3) that repression is of aetiological significance in slips of the tongue; and so on. The first has been supported. The others have not, as yet, although, it must be noted, they have not been refuted.

I hope now that is clear why I have argued that Grünbaum's devastating philosophical analysis of the clinical basis of Freud's theory of repression is not so destructive to psychoanalytic theory as a whole as perhaps Grünbaum would have us believe. A different, but certainly viable, approach to the validation of psychoanalytic theory reveals that something of value can be salvaged. I also hope that I have demonstrated that the flaws in the *Foundations of Psychoanalysis* arise not from any philosophical weakness but from straying into psychology where naturally Grünbaum is less at home. Indeed it is ironic that Grünbaum should acknowledge the help of Eysenck and Rachmann in getting to grips with understanding the research in psychotherapy, for these are two of the most relentless and bitter critics of psychoanalysis. It is as if a naive visitor to Heaven were to have Lucifer as guide.

Grünbaum, in his conclusions, states that psychoanalysis is hardly well at least in so far as its clinical foundations are concerned. Nor is there a favourable verdict from such experimental findings as were cited in the text. In brief, for the reasons that I have put forward in this article, Grünbaum has not made his case. He has demonstrated that the clinical basis of repression is ill-founded, but this is not the whole of psychoanalysis. Furthermore, his appraisal of the experimental evidence cannot be taken seriously without presentation of detailed reasons for rejection or acceptance of studies which Grünbaum does not offer. Despite this it is a brilliant and powerful analysis of the philosophical basis of repression, one which none interested in psychoanalysis can ignore.

PAUL KLINE
Department of Psychology, University of Exeter

REFERENCES

EYSENCK, H. J. and WILSON, G. D. [1973]: *The Experimental Study of Freudian Theories.* Methuen.

FARRELL, B. A. [1961]: 'Can Psychoanalysis be Refuted?' *Inquiry*, **4**(1), 16–36.

FARRELL, B. A. [1981]: *The Standing of Psychoanalysis.* Oxford University Press.

FISHER, S. and GREENBERG, P. R. [1977]: *The Scientific Credibility of Freud's Theories and Therapy.* Harvester Press.

FREUD, S. and BREUER, J. [1893]: *On the psychical mechanism of hysterical phenomena: preliminary communication.*

FREUD, S. [1917]: *Introductory lectures on psychoanalysis.*

FREUD, S. [1925]: *An autobiographical study.*

FREUD, S. [1937]: *Construction in analysis.*

KLINE, P. [1972]: *Fact and Fantasy in Freudian Theory* (1st edn). Methuen.

KLINE, P. [1981]: *Fact and Fantasy in Freudian Theory* (2nd edn). Methuen.

KLINE, P. [1983]: 'Meta-analysis, Measurement and Methodological Problems in the Study of Psychotherapy', *Behaviour and Brain Sciences*, **6**, 296–7.

MARTIN, M. [1964]: 'Mr. Farrell and the refutability of psychoanalysis', *Inquiry*, **7**, 80–98.

MAYER-GROSS, W., SLATER, E. and ROTH, M. [1967]: *Clinical Psychiatry.* Cassell.

SEARS, R. R. [1943]: *Survey of Objective Studies of Psychoanalytic Concepts* (Bulletin 51). Social Science Research Council.

SMITH, M. L. and GLASS, G. V. [1977]: 'Meta-Analysis of Psychotherapy Outcome Studies', *American Psychologist*, **32**, 752–60.

SMITH, M. L., GLASS, G. V. and MILLER, T. I. [1980]: *The Benefits of Psychotherapy.* John Hopkins University Press.

ZAMANSKY, H. S. [1958]: 'An investigation of the psychoanalytic theory of paranoid delusions', *J. Pers.*, **26**, 410–25.

Brit. J. Phil. Sci. 38 (1987), 117–139 *Printed in Great Britain*

Reviews

HOOKWAY, CHRISTOPHER [1985]: *Peirce*. Routledge & Kegan Paul. *xii* +301 pp. (Hardback £25.00). (ISBN 0–7100–9715–8).

This is a competent book on a difficult subject. It surveys many aspects of Peirce's thought, and for the most part clarifies rather than obscures its subject. Its only weakness is that it tries to make Peirce serve the needs of one strand of contemporary Anglo-American philosophy.

Charles Peirce is generally regarded as the foremost philosopher to have lived in the Americas, and studies of his ideas have grown from cottage production to major industrial strength. There is a philosophical society devoted to protecting his repute; an institute and a newsletter exclusively promulgating his ideas; and a magazine whose major function is to provide a forum for Peirce scholarship.

At the same time the interpretation of Peirce's work is in disarray. Early in his adulthood, he was forced from the American university system, and his career and life were disrupted for forty years thereafter. He never completed a book-length treatise, and much of his writing is still in masses of fragmentary manuscripts. Beginning in the 1920s Harvard published eight volumes of Peirce's 'collected papers', but the editorial procedures left much to be desired. Important writing was not published, and much that was published was difficult to construe—material written at different times on different topics was juxtaposed, material written at the same time on one topic was separated; bits of individual essays were published, bits were left unpublished; the writing was ordered in a way that now seems arbitrary. Finally, Peirce's style throughout was peculiar—baroque, ellipitical, alternatively self-effacing and magisterially egoistic.

The result is a body of scholarship that is almost as chaotic as Peirce's unpublished papers, with widely varying accounts proliferating in the literature. A new edition of Peirce's writing is just now appearing, and will make available a larger selection of his work, organized around chronological principles. The completion of this edition, along with easier access to a microfilm edition of what remains and a catalogue of the papers, will, I believe, considerably reduce the chaos. But until that time Peirce scholarship lacks a definitive base.

The major virtue of Hookway's *Peirce* is that it presents a fair outline of much of Peirce's thought and makes a comprehensive attempt at discussing the secondary scholarship. Nonetheless, Hookway weakens his presentation by inevitably tailoring it to what he perceives to be the needs of an excessively narrow readership.

Peirce is part of a series entitled 'The Arguments of the Philosophers', and it may be that this series is designed to mine the still usable arguments of past thinkers and then present them in mint condition for the use of philosophers at London or Berkeley. Whether or not this is so, Hookway often elucidates Peirce in this way.

The best part of this book, I believe, is the overview of Peirce's early thought, from which Hookway branches out to study selected topics systematically. But Hookway is a little dismissive of the young Peirce and, he says, does not detail his arguments because many of them are worked out in 'a sketchy fashion'. Perhaps; but it may be that Peirce did not intend to produce arguments for 'the arguments of the philosophers' series and that *his* details might suggest a different way of philosophizing.

Hookway knows a lot about philosophers Peirce didn't know about. At various times to illuminate Peirce, Hookway compares his ideas to those of Anscombe, Goodman, Quine, Putnam, Rawls, Rorty, and Sellars. At two points he gives Peirce high praise by saying that like *Frege* 'he recognizably inhabits our philosophical world'; Frege's notions, says Hookway, are likely to be familiar to many readers, and allow us to elaborate Peirce's. In examining a very famous 1870 paper of Peirce's, Hookway writes that it will be 'surprising' to the 'modern ear' because its realism lacks what *Dummett* supposes to be the distinctive characteristic of realism. Although Hookway thinks Peirce's metaphysics queer enough to call them 'bizarre' in two places, Peirce's views are 'less wildly speculative—and less alien to the styles of modern philosophy' than they may appear. God save us from this condescension of the present!

On the other hand, Hookway spends little time with the thinkers who were Peirce's contemporaries, and certainly gives the impression that he has not read much of what Peirce read. The views of Hamilton, Bradley, and Bain are scarcely touched. Asa Gray becomes Asa Gay. About James and Dewey Hookway only remarks that Peirce's pragmatism is not their 'irrationalist doctrine'. Hookway notes that Peirce himself considered his views to be most similar to those of Josiah Royce, whose 1885 argument for absolute idealism was as influential in its time as Quine's 'Two Dogmas' was sixty-five years later. Hookway muddles the date of Royce's work, and in a few sentences finds it 'very obscure'. In the one serious and pervasive distortion Hookway neglects to examine the contours of F. E. Abbott's arguments about nominalism and realism, and consequently leaves his own discussion of Peirce's realism unclear.

In treating Peirce's important predecessors, Hookway is better. He says that he lacks the space to discuss Peirce's knowledge of medieval thought, perhaps the most serious omission in the book; and he has little to say of Berkeley or James Mill, two historical foes of Peirce. But Hookway does have an outstanding discussion of the connection between Peirce and Kant—in whose steps Peirce walked—which provides the main organizing theme of the book. Thank God Kant's stock has gone up in Oxford and Cambridge in the last twenty years!

Long ago Peirce was claimed by logical positivism, and more recently by analytic philosophers, phenomenologists, process thought, and semioticians. It is not surprising that post-analytic philosophy in the style of Putnam and Dummett finds truth in Peirce. But it is likely that only philosophers with that peculiar range of interests will find this book of use.

The reader, I hope, will forgive the testy nature of some of my remarks. For *Peirce* is, as I have said, a solid survey of many of Peirce's doctrines. Nonetheless, Hookway also displays a remarkable insensitivity to history and an alarming philosophical parochialism.

BRUCE KUKLICK
University of Pennsylvania

ROSENBERG, ALEXANDER [1985]: *The Structure of Biological Science.* Cambridge University Press. xi+281 pp.

Rosenberg's book serves two purposes: it is an introduction to the philosophy of biology and an attempt to argue for a compromise between autonomism and antiautonimism. Rosenberg more than achieves the first goal; in fact, his book may be the best introduction to the philosophy of biology yet written. He also makes a noble attempt to achieve the second goal, but he fails to show how his view differs from autonomism and presents an unconvincing argument for the possibility of reducing biology to physics and chemistry.

Rosenberg tackles smaller problems in the philosophy of biology in order to defend his answer to the large question of the autonomy of biology. He reviews debates concerning teleology, the reduction of Mendelian genetics to molecular genetics, and the nature of evolutionary theory (including discussions of the concepts of 'species', 'adaptation', and 'fitness'). He also adds a final section on metaphors in biology.

The first issue Rosenberg addresses is the status of teleology. Autonomists argue that teleological explanations make biology an autonomous science because they cannot be reformulated as nonteleological explanations. Antiautonomists counter that teleological explanations can be reduced to nonteleological explanations, or if they cannot be reduced, then they should be eliminated from biology. Rosenberg tries to avoid taking either of these extreme views: he argues that even though reduction is in principle possible, biologists must adopt a functional perspective because of the limits of human reason.

Rosenberg endorses materialism—all life is nothing but matter in motion. He also argues that it is in principle possible to explain everything in terms of physical laws if we assume nature obeys laws as deterministic as those of quantum mechanics and that there is a finite amount of matter in the universe. According to Rosenberg, only a finite number of things can happen in the universe because there can only be a finite amount of arrangements of matter in the universe. Since only a finite amount of arrangements of matter are possible, teleological processes can be physically realized in only a finite number of different ways. Therefore, "In principle, there will always be a nonteleological description of the internal components of a system that together with the causal laws about these components determine the persistence and plasticity of its goal-directed behaviour" (p. 63).

But the reducibility of biology has little practical import because we cannot formulate and understand theories which explain life nonteleologically. When we try to make this reduction, we find that our theories are too complicated and our explanations have too many disjuncts in their premises to allow us to make inferences. A Laplacian genius might be able to understand all the mechanisms that lead from atoms to organisms, and we can provide nonteleological explanations for many biological processes, but we must accept teleological explanations in biology.

The practical irreducibility of teleological explanations gives biology practical autonomy but not autonomy *simpliciter*: it is still in principle possible to apply the methods of explanation and discovery used in physics and chemistry to biology, but biologists may find it more useful to employ their own teleological explanations in many cases. To illustrate this point, Rosenberg considers a teleological explanation for ATP production. Mammals produce ATP. A common biological explanation for this phenomenon is that mammals have the goal of producing ATP. Supposing that mammals have such a goal has allowed biologists to construct extremely useful explanations of ATP production. In contrast, a nonteleological account of ATP production must enumerate the three different processes which produce ATP (glycolysis, the Krebs cycle, and oxidative phosphorylation), as well as the various subsystems and enzymes within these processes which contribute to ATP production. Since many different processes may generate ATP (its production is causally overdetermined), the premises of the nonteleological explanation must contain many disjuncts (*e.g.*, the mammal performed either oxidative phosphorilation or glycolysis, or the Krebs cycle). As biologists analyze more complex systems, it becomes harder to enumerate all of the disjuncts and they find that a nonteleological account becomes "unmanagably long, baroque, and without further theoretical and predictive payoff beyond the functional explanation" (p. 62).

In order to reconcile the need for teleological explanations with a commitment to reductionism, Rosenberg maintains that many biological properties, explanations, laws, and theories supervene on the laws of physics. He argues for the supervenience of Mendelian genetics on molecular genetics: "Mendelian objects and properties supervene [on] molecular ones: The latter fix the former ones completely, even though they cannot be connected to them in any manageable way" (p. 113). Similarly, Rosenberg maintains that the evolutionary fitness of organisms supervenes on their physical properties.

Although Rosenberg carefully describes the supervenience relation and tries not to endow biology with full-fledged autonomy, unless he provides us with a method for reducing biology to physics and chemistry, his "merely practical" autonomy collapses into autonomy *simpliciter*. Rosenberg gives an argument for reduction as an abstract, logical possibility and urges that we could list all of the possible physical realizations of biological goals and draw the weak conclusion that one of them must obtain. This "reduction" presupposes that we know all of the possible physical realizations. But how can we discover all of these possible realizations? By chance? Unless Rosenberg discovers a method for deriving all possible physical realizations of biological processes from their teleological descriptions, we cannot make a reduction because we can never be sure we have enumerated all of the possible physical realizations. Even if we obtain a method for reducing teleological explanations, it may never be possible for human beings to make this reduction because it exceeds our computational capabilities and our abilities to manipulate symbols. If we can't make this reduction, in what sense is reduction possible? As far as human epistemeology is concerned, biology is an autonomous science.

Rosengerg's argument for the logical possibility of reduction rests on disputable assumptions. Besides appealing to a weak determinism, he assumes a controversial view of space-time. Even if we grant him that there is only a finite amount of currently existing matter, an infinite amount of spatio-temporal arrangements of this matter would still be possible, unless we adopt a finite geometry. If there are an infinite number of possible arrangements of matter, then there can be an infinite number of physical realizations of biological functions. Hence, nonteleological reformulations of teleological explanations could have an infite number of disjuncts in their premises, and reduction would be logically impossible. Even Laplacian geniuses would have to use teleology.

Rosenberg fails to provide a compelling argument for his philosophy of biology, but he succeeds in writing an excellent introduction to the philosophy of biology. He focuses on problems of immediate concern in biology and gives a thorough treatment of the philosophical issues in biology. His book contains more biology than previous introductions to the philosophy of biology, which gives the reader an excellent opportunity to learn a significant amount of contemporary biology. *The Structure of Biological Science* is an impressive book, and both biologists and philosophers will benefit from reading it.

DAVID B. RESNIK
The University of North Carolina at Chapel Hill

I am indebted to William G. Lycan, Geoffrey Sayre McCord, and Michael D. Resnik for reading and commenting on this review.

DENBIGH, KENNETH G. [1981]: *Three Concepts of Time*. Springer-Verlag. *viii* + 180 pp. $17.80.

It is comparatively rare to find an author writing on the highly recalcitrant topics of the physics and metaphysics of time who combines genuine philosophic insight with a first hand knowledge of contemporary physics as does the author of this volume. The book has two parts, the first sets up the problem and the second part being a detailed study of temporal processes, provides the author's solution and his argument for it. The theme of Part I is that there are essentially three distinct and at first sight irreconcilable concepts of time *viz*: those temporal concepts based upon conscious awareness (the present, past and future), the concept of time as employed in fundamental physics (the time symmetric continuum of temporal instants) and the asymmetric, anisotropic temporal continuum employed in study of irreversible natural processes. The problem to be solved is how are these three conceptual structures interrelated and to what extent if any can they be integrated into a coherent whole? Denbigh's thesis is that they can be so integrated provided that they are treated as 'different levels of description' forming a hierarchy. Part II of the volume provides the analysis of, and argument for this claim by way of a systematic discussion of temporal processes. The argument of Part II is very wide ranging and the author draws upon material not only from the standard areas of the physics and philosophy of time but also treats of determinism, chance, reductionism, information theory, the theory of organised systems, neurobiology and the philosophy of mind.

Now Denbigh is certainly right to distinguish the three sets of temporal concepts but to this reviewer at least there can be no hope of integrating all of them into one consistent conceptual structure no matter what standpoint is adopted. The reason is that *if* one takes the ordinary common-sense concepts of past, present and future as presupposing the reality of tense, that is of presupposing that everything in time has successively the real properties of being future, of being present and of being past and that this accounts for change, then that presupposition rules out consistently uniting the ordinary concepts with the temporal concepts of science. Now Denbigh argues that whether the antecedent of the above conditional holds or not is an unanswerable question (pp. 66–7) and that this shows that progress in the theory of time can only come from adopting a different perspective. He writes

> . . . 'time' should not be conceived as some sort of *existent* since it could hardly be said that time exists in time. Thus it is not a physical entity having its own intrinsic properties—i.e. determinate characteristics which are available to be discovered. It is rather an abstract relational concept. As such its *attributes*, and also the number of them, depend on the level of the particular phenomena where specifically temporal relations need to be established. *It is a question of different levels of description.* (p. 67)

When this is done it is his claim that we can obtain from the three disparate concepts a consistent conceptual structure. In order to assess the author's claim we need first to see what he takes to be the core of the three temporal concepts and then examine how the 'new perspective' gives a unifying consistent treatment of them.

Fundamental to the account of time in conscious awareness and time in physics developed in Chapters 2 and 3 are two senses of objectivity of judgement (p. 11). The first sense (objectivity$_1$) is characterised by appeal to the possibility of intersubjective agreement, while the second sense (objectivity$_2$) 'relates to statements about things which can be held to exist, or about events which can be held to occur, quite independently of man's thoughts or emotions, or his presence in the World.' Objectivity$_1$ is intended to ground use of the ordinary temporal concepts of conscious awareness and objectivity$_2$ to characterise temporal concepts as used in physics. Now while these characterisations of objectivity are not intended by the author to be exhaustive (and certainly not exclusive) it is worth noting just how far short they fall of capturing the notion of objectivity of judgement. For example one may well obtain inter-subjective agreement that some event was funny or absurd without necessarily being committed to the view that some objective judgement about the situation had been arrived at. Rather the required sense of 'objectivity' must have much more to do with the claim that the judgement arrived at is independent of the method of investigation employed on a given occasion. This point is not intended as a specific criticism of the author's account but rather as a general point about the difficulty of describing just what is claimed when it is said that a given judgement is objective, whether that judgement occurs in common sense or within physical science.

A central claim of the author is for the logical and epistemological priority of the temporal concepts of conscious awareness, in particular of 'the present' (pp. 14, 27, 39, 125). Chapter 2 is devoted to constructing the intuitive common-sense concept of time as a single ordered sequence (with the order type of the natural numbers) out of presents or moments, a single ordered sequence being obtained out of the sequence of presents for each individual by appeal to public agreement (objectivity$_1$) as to events occurring at the same moment or present. The ordering relation of 'later than' is on this account rooted in the faculty of memory. Chapter 3 gives an excellent exposition of the temporal *continuum* of physics which 'makes no reference whatsoever to "past", "present" or "future"' (p. 27). Given that the basic laws of physics are time-symmetric and that the physical temporal continuum is orientable the question arises as to how the anisotropy (the fact that physical processes always occur in only one of the two distinct time senses) of the continuum is to be explained? Denbigh employs Schrodinger's form of the second law of thermodynamics (pp. 35–39) to the effect that for any two entropy creating systems A and B and any two instants t, t' (denoting the entropy of system A at t by S_{A_t} and similarly for B

and t')

$$(S_{A_t} - S_{A_{t'}})(S_{B_t} - S_{B_{t'}}) \geqslant 0.$$

But this is of course the reversible form of the second law expressing only the parallelism of all entropy changes, and so will not serve to establish the anisotropy of time unless it be coupled with the claim of the existence of irreversible processes. Given that existential claim be satisfied then if the entropy change for a system undergoing an irreversible process in the interval (t, t') is negative (positive) then it will be negative (positive) for all other physical systems undergoing irreversible change of state over that interval. In that sense all physical processes occur in one time *sense*, the time sense that renders the entropy change over that interval to be negative (positive). But this does not establish as such the sense of 'later than'. That is only established once it is found 'that the *higher* entropy states are those which occur later in my consciousness' (p. 36). That is, what is constitutive of the anisotropy of time is the fact that physical processes are (de facto) *irreversible*. A criterion for this irreversibility is the change in a function of state variables of physical systems, the entropy. All changes of entropy for adiabatically isolated systems occur in the same sense, either all increasing or all decreasing. If we now assign to the higher entropy states the later times, then we have in the increasing entropy a physical criterion for later states, but we have in no sense given a definition of 'later than'. These points are well worth making and Denbigh's account of them is very clear.

Now while this reviewer has no quarrel with any of the above, nor indeed with the author's lucid treatment of the thermodynamic (in the sense of statistical mechanics) origin of *all* irreversibility and the Reichenbach–Grünbaum hypothesis of branch systems (pp. 93–130) where there is a considerable lacuna in the author's argument lies just in the connection he claims obtains between the concepts of temporal order occurring in conscious awareness and those occurring in physics. The author is a relationist about time (pp. 3–4, 11), so the question naturally arises as to the relations among what constitute temporal order? It cannot be at bottom merely the relations among *experiences* for the author clearly believes that there are temporal relations among events, independently of anyone's perception of them. The author however eschews the causal theory of time (pp. 31–2) which would be the natural candidate for connecting the temporal order of events with the temporal order of our experiences of them (in conscious awareness, perceptions too being events). But we are told that the logically and epistemologically prior notion is that of the temporal order of conscious awareness, but then what is it conscious awareness of? Of course, if one were an absolutist about time, one might say that we simply possessed a faculty for directly apprehending in conscious awareness the passage of time (perhaps by analogy with the epistemology of Platonism in mathematics) but the author given his relationist view, rightly rejects this possibility as incoherent (p. 94). But the question then is immediate: of what are we consciously aware when we perceive temporal order? Following many, my own suggestion would be to answer of the causal relations among events. But this gives both the logical and epistemological priority to those relations among events which we perceive as their temporal order. But claims about such relations obtaining among events would be objective$_2$ claims and so priority would go there and not as the author would have it to the temporal concepts of conscious awareness (the objective$_1$ sense of time).

A second and related objection concerns the author's treatment of the classical division between A—theorists (who regard the essence of time as change and assert that change can only be explained by postulating real non-relational tensed properties) and B—theorists (who claim that a complete account of temporal phenomena can be obtained using the tenseless relations, 'simultaneous with' and 'later than') concerning the reality of tense. Clearly the author feels that this debate has not been helpful, and indeed has perhaps been obscurantist. As to the question of the reality of 'the present' he remarks: 'Since everything so far seems to imply that this is an unanswerable question it is desirable to attempt seeing the issue in an entirely different perspective' (p. 67). The different perspective which is proposed is that of the 'naturalistic viewpoint' (pp. 64–68). The trouble with this is, as elsewhere in epistemology, that it merely pushes the problem one back. We can all agree that man is part of the natural world and therefore subject to its laws and to all the evolutionary constraints on biological phenomena. But still the question will remain as to whether those events, occurring in the brain and related to memory and perceptions, do or do not like all other events sustain the non-relational property of 'presentness' *etc.* Essentially the question as to the existence of tense (as opposed to its ineliminability from discourse) and its centrality in the philosophy of time cannot be outmanoeuvred by adopting a 'naturalistic' standpoint. This of course is not to say that some progress cannot be made by adopting that standpoint, but at least one central issue will not go away.

The second half of the volume devoted to temporal processes contains a substantial discussion of irreversibility (Chapter 6) which is excellent and in Chapters 7 and 8 a fine study of irreversible process in biology and of the biological evolution of consciousness. This latter includes a detailed discussion of the interrelated concepts of chance, organisation and negentropy (pp. 136–52), which certainly clears up a number of conceptual confusions to which the literature in this area is prone, particularly among the concepts 'information', 'degree of organisation' and 'negentropy'. To this reviewer however there is one unsatisfactory aspect to Part II which is the author's treatment of determinism and chance (Chapter 5). The author distinguishes as is traditional two forms of the deterministic thesis, predictive determinism and ontological determinism. Predictably he has no sympathy for ontological determinism, regarding the only reasonable form of determinism as the predictive kind. (Predictive determinism he takes to be the thesis that an 'event *E* becomes more and more accurately predictable the shorter is the temporal interval between the moment of prediction and the moment t_E at which the event occurs.' (p. 72)). However the situation is surely just the other way. Predictive determinism, either in the form that *all* events are predictable according to the methods of science or in Denbigh's temporal version, is an extraordinarily strong thesis instantiated by none but the simplest examples of classical mechanics (like two particle systems) whereas ontological determinism is instantiated by all of classical mechanics and relativity theory including of course classical statistical mechanics. Indeed it is surely a precondition for setting up classical statistical mechanics that the underlying motion (the trajectories) of the elements of the ensemble be deterministic, precisely in the sense that no two such trajectories may intersect. But of course such underlying motions are manifestly unpredictable, even in Denbigh's more interesting version of predictive determinism. Now of course it is true that 'ontological' determinism is very difficult to formulate, if we really want to rid it of all epistemological taint, but a form of the deterministic thesis which does not rely on predictability can be obtained via the notion of a deterministic theory. Roughly a theory is deterministic when any two systems which are models of the theory, having *identical* states at one time must have *identical* states at all times. It is precisely in this sense that classical statistical mechanics is deterministic in the underlying mechanics. Just because a system is deterministic it certainly does not follow that its states cannot sustain a non trivial probability measure. For example an ergodic system (one in which time and phase averages are identical) like that of a system of rigid billiard balls moving within a parallelopiped certainly sustains a probability measure defined on its state-space, yet that motion is entirely deterministic in the above sense. Nor more importantly is it the case that deterministic systems must always generate randomness in the values of state variables by having that randomness injected in at the level of initial conditions for as long as the functions connecting the values of state variables are not effectively computable a sequence of such values may well be deterministic yet random in the sense of Von Mises-Church.

This volume provides a very full introduction to the main issues in the philosophy of time. All of the great questions about time; what is the order-type of the set of temporal instants, the conventionality or otherwise of temporal congruence, the beginning and end of time, its anisotropy and its topology are all treated to some extent or other. What the author has to say about them is insightful and interesting. The volume is manageable and pellucidly clear. It deserves to be very widely read.

PETER CLARK
The University of St. Andrews

RUBEN, D.-H. [1985]: *The Metaphysics of the Social World.* Routledge and Kegan Paul. $x + 189$ pp. £14.95.

1 Analytic philosophy has tended to concentrate on the methodology of the social sciences, paying relatively little attention to problems about the ontology of the social world. Perhaps the thought that underlies this selective interest is that it is the social scientist's task to tell us what the social world is like, the philosopher's task is to rationally reconstruct the activity of the social scientist. But just as there are problems about the physical world that are the philosopher's business (*e.g.*, What is a law of nature?) so there are problems about the social world that must be treated of analytically, rather than by empirical methods. Answers to these questions may have implications for methodology, but this is not their only justification. The question, What makes a relation a social relation? would be of philosophical interest even if it demonstrably shed no light on what social scientists are up to.

Anyone sympathetic to this view ought, I think, to welcome David-Hillel Ruben's new book. He deploys the methods of analytical metaphysics in defence of an holistic view of the social world. He defends the irreducibility of the social to the individual at both an ontological and an explanatory level. He believes that there are distinctively social objects, and distinctively social properties (chapters 1 and 3). He believes that satisfactory explanations of social phenomena will inevitably make reference to both these kinds of entities (chapter 4).

While Ruben's position is holistic in that he thinks that social entities are *sui generis*, it carries no commitment to the idea that the individual and the social are related as parts and wholes; he offers good arguments for thinking that they are not (chapter 2). Ruben is best described, I think, as a *dualist* about the social. The analogy with the mental that this invites is, as we shall see, an important one.

I shall say more about Ruben's arguments for the irreducibility of social objects below. He has two arguments for the irreducibility of social properties; their multiple realizability in terms of individual properties, and their essential occurrence in ascriptions of propositional attitudes. His argument against explanatory individualism is roughly this: statements about an individual's beliefs require, where those beliefs are rational, explanation in terms of the states of affairs referred to in the content clauses of the corresponding attitude ascriptions. But these states of affairs will sometimes be social and, by the arguments of chapters 1 and 3, they will be irreducible. So we cannot expect explanatory chains to discharge their commitments to the social after a certain number of backward steps (pp. 168–71).

I shall not summarize the book's contents beyond what I have said so far. Instead I shall develop a couple of its themes. The first point I want to make is critical of Ruben's anti-individualism. The second supports his characterization of a social property. Both my arguments draw upon the work of David Lewis.

2 Ruben argues in his first chapter against an individualistic account of social substances. France, Ealing and the Red Cross are examples of social substances, while voting and purchasing are examples of social properties. Ruben argues first of all that a social substance like France cannot be identified with a set of individuals—the set of people who were, are or will be French, for example. His main argument here is that such an identification cannot make sense of counterfactual claims like 'Bismark might have been French', for sets are extensional and preserve their membership across possible worlds (p. 17). Now it is clear that Ruben treats names of social substances as rigid designators. One would not want to argue as follows: 'the King of France cannot be an individual, for if it was it could not have been any individual other than the one it is'. If 'France' designates non-rigidly, it is open to us to say that 'France' denotes a set of individuals, but distinct sets in distinct possible worlds, and the modal objection collapses.[1]

Now Ruben also argues that France cannot be identified with any group of individuals, where a group is specified not in terms of set membership but rather in terms of possession of a common property. This would be one way of reading 'France' as non-rigid, for we would have

$$\text{France} = \{x : Px\}, \tag{1}$$

where P is the favoured property; a property that takes different extensions in different worlds. Now Ruben argues that the only reasonable candidates for P are such that their instantiation presupposes the existence of France itself; they fail, therefore, to provide definitions of 'France' (pp. 26–44). I do not quarrel with this. But not all interpretations of 'France' as non-rigid take the form of (1). Let us look at one.

David Lewis has suggested that we may identify mental states with physical states in the following way.[2] We are to characterize mental states in terms of their causal roles; it will then turn out that the occupants of these causal roles are brain states. So while the concept of pain is not the concept of a physical state, everything that is a pain is, actually, a physical state. Now Lewis notes a threatening circularity here, because the specification of the causal role of a mental state may involve reference to other mental states

---

[1] Here I assume a distinction—rigid/non-rigid—that I am inclined to reject. I prefer to draw the distinction in terms of office-denoting terms and individual-denoting terms, and to construe claims like 'Louis Capet is the King of France' as stating that an individual occupies an office. But for present purposes this is not an important issue; the individualism that I defend here can be construed on the office model as saying that social-office denoting terms ('France', *etc.*) denote offices that are always occupied by sets (or structures) of individuals. For more on the general semantical issues at stake here and for further references, see Currie and Oddie [1980].

[2] See Lewis [1970] and [1972].

(pain causes anxiety, for instance). We avoid the circularity by defining mental states all at once, so to speak. Let $T[t_1, \ldots, t_n]$ be 'folk psychology'—the theory that sums up our everyday knowledge about mental states and their causal roles, where the t_i's are names of mental states. Then the 'Ramsey Sentence' of T is

$$\exists x_1, \ldots, \exists x_n T[x_1, \ldots, x_n].$$

Suppose that t_1 is 'pain'. Then we can define

$$\text{Pain} = \iota x_1 \exists x_2, \ldots, \exists x_n T[x_1, \ldots, x_n], \tag{2}$$

and so on for all the other mental states referred to in T.

So pain is defined as that unique thing such that . . . , where the blank is filled in by folk psychology. Mental states are defined in terms of their relations to one another (and to behaviour), without circularity.

In the same way, let T be our geo-political theory, telling us all we want to know about the countries of the world and the history of their relations to one another. T will contain terms $t_1, \ldots, t_n$, naming countries. Let t_1 be 'France'. Then we define France as in (2), substituting 'France' for 'pain', and so on for all the other countries referred to in T.

T specifies a number of geo-political roles. The individualist's claim will then be that the occupants of these roles are, as a matter of fact, sets of individuals; distinct such sets in distinct possible worlds, no doubt.

But, it will be objected, if we define countries in this way, surely it will turn out that France could not have had a history different from the history it does have, for countries like France will exist only in worlds where T is true. And Ruben rightly points out that claims like 'France was defeated by Prussia in 1870' are merely contingent (p. 37). Lewis has a suggestion that will help us here. Consider the following story. The police are wondering about who has committed the murders in Whitechapel. They introduce the expression 'Jack the Ripper' to designate the person who . . . , where the dots are filled in by a lengthy description distilled from the available evidence. Then Jack the Ripper is the person, whoever he is, who fits the description. But of course no one may fit the description perfectly. It may be part of the description that Jack killed six women. But perhaps only five were killed by the same person; the sixth was a copycat crime. We would not want to say in that case that nobody was Jack the Ripper. The perpetrator of five of the killings would surely fit the bill, because he is the closest thing there is to a satisfier of the description, and he comes close enough. If the police description is hopelessly wrong, and the killings were all done by different people, then no one would be close enough to fit the bill.

Now I think that our identification of countries may be on similar lines. That is, something will be France in a world where nothing quite fits the relevant parts of the description in T, our geo-political theory, as long as it is the closest thing there is to fitting the description, and it is reasonably close.

Thus we may say that 'France was defeated by Prussia in 1870' is contingent, since there is a world in which 'France' refers to an entity (set of individuals, the individualist will say) that fails to satisfy that particular bit of *T*, but satisfies a good deal else of it. The less of the description the object satisfies, the less happy we will be about calling it 'France'.[1] Different parts of the description summarized in *T* may get different weights. Quite marked counterfactual deviations in the recent history of France would probably permit reidentification, as long as the early history was close to the early part of France's actual history. No philosophy can hope to specify the exact principles governing these weighting decisions, but it seems intuitively correct to say that there will come a point where the counterfactual deviations are so marked that reidentification is no longer possible. The essential point is that identifying a social institution in a counterfactual situation is not a matter of tracking a particular object across possible worlds, but of deciding what object best fits a certain institutional role. And this seems right: for how else do we identify institutions in the first place other than by citing the roles they play?[2] (In this respect Lewis's theory works better for institutions than for mental states, at least some of which we identify in terms of their phenomenal qualities.)

So far I have been defending a view that is compatible with the claim that social substances are sets of individuals. But there may still be good reasons for saying that they are not sets of individuals. Is a car the set of its mechanical parts? Not when the parts are distributed around the globe. The car is more plausibly identified with the set of parts as functionally related; a structure in other words. Perhaps we should say then that the occupants of geo-political roles are structures of individuals as socially related. As Ruben notes, Popper is an individualist who is prepared to countenance unreduced social properties (p. 175). Unless the Popperian individualist has Leibnizian scruples about relations, he or she will probably admit social relations between individuals as well. All that is denied is the existence of unreduced social substances. Thus the Popperian ontology of the social has similarities with the ontology of the token-token materialist who denies that there are distinctively mental substances but sees no prospect for a reduction of mental properties. The consequences of such a view of the social world have not been well developed, especially in relation to problems about explanation, but it does seem to be the most viable form of individualism that we have.

---

[1] As David Lewis pointed out to me, we need to distinguish *de re* and *de dicto* counterfactuals here. 'France might not have been defeated' is unproblematically true on the *de re* reading that also makes 'France might not have been France' true. It is the *de dicto* reading that we are trying to capture here.

[2] This approach may be useful in solving some of the problems concerning terms like 'the Renaissance' whose reference is to some degree indeterminate, and variable from one theoretical context to another. For an interesting discussion see Stainaker [1967]. This is the only article I know of in which problems of reference in the social sciences have been seriously discussed.

Thus far an individualistic response to Ruben's dualism. But one argument against such a dualism should be jettisoned: that dualism attributes modes of independent action to social entities. One can accept, on the contrary, that social objects are irreducible to individualistic ones, and hold that social facts are entirely supervenient upon individualistic ones. Such a supervenience thesis would, I think, be a natural adjunct to Ruben's metaphysics.[1]

3 Ruben gives us an explication of what it is for something to be a social property: a property is social if its instantiation entails the holding of a certain kind of relation between individuals. The most direct form of this relation (there are other, more distant forms of it) holds between *A* and *B* when it is common knowledge between *A* and *B* that (1) *A*'s awareness of *B*'s expectations about *A*'s behaviour influence *A*'s behaviour and (2) similarly for *B*'s awareness of . . . (pp. 107–14). Let us call such a relation an *S*-relation. Ruben does not give a detailed argument as to why *S*-relations are the mark of the social, and it may be that there are allowable senses of 'social' that are weaker than this. Nevertheless, I think that Ruben has delimited a very important sense of the social; I should like to call it the 'full' sense, for the following reason.

David Lewis has given an account of what it is for something to be a convention.[2] A regularity *R* for a population *P* is a convention *iff* (*i*) everyone in *P* conforms to *R*; (*ii*) everyone expects everyone else to conform to *R*; (*iii*) everyone prefers to conform to *R* on condition that the others do, since *R* solves a coordination problem for members of *P*. Now Ruben's strategy is to consider instead two populations isolated from one another, and then gradually to build up connections between them to the point where the connections constitute an *S*-relation. At this point we have, he claims, a social relation. I shall try to show that this is right. For the sake of simplicity I shall consider two individuals isolated from one another. I do not think that this simplifying assumption will affect the argument I am going to give.

Suppose that *A* and *B* are individuals who live on opposite sides of a wide chasm. So far there is no contact between them. Now it might be to their mutual advantage if each can let the other know that a storm is approaching (supposing them to have different perspectives on the weather). They have a coordination problem, and they need to establish a convention whereby the observable behaviour of the one is taken by the other as meaning that a storm is on the way. *A* sees a storm approaching and waves his arms at a place where he knows that *B* can see him. *B* then catches on (perhaps after a number of trials) that *A* is signalling a storm, and it will be to his advantage to use the same behaviour when he sees a storm approaching from his side, for this will reinforce *A*'s use of the regularity. As Lewis points out, following the regularity

---

[1] See Currie [1984].

[2] See Lewis [1969], especially p. 42. I follow Bennett [1976] in his application of Lewis's account to the coordination of utterances.

(*i*) arm wavers intend observers to believe that a storm is approaching;
(*ii*) observers of arm wavers come to believe that a storm is approaching,

will involve a Gricean intention if the regularity is a convention.[1] But in order to establish such a convention I think that *A* and *B* must

(1) have expectations about each other's behaviour;
(2) have common knowledge of (1);
(3) have (2) as a reason for behaving;
(4) have common knowledge of (3).

Thus *A* initiates the regularity hoping that *B* will join in. So he must think that *B*'s behaviour can be influenced by his own. But he will not think that *B*'s behaviour can be merely *caused* by his own; he must think that *B*'s behaviour can be influenced by *B*'s awareness that *A* expects something of him. Similarly for *B*, if he is to initiate the regularity. But this is to say just that conditions (1)–(4) are necessary and sufficient for the establishment of a convention and hence sufficient for the operation of the Gricean mechanism. But (1)–(4) amount to Ruben's conditions on an *S*-relation. So if you think, as I do, that the establishment of conventions and the operation of Gricean communication are essential aspects of the social in the full sense, you will think, as I do, that Ruben's conditions are the right ones.

4 Reading this book one gets a strong sense that the range of options in social philosophy is richer, more complex, and more dependent upon abstract metaphysical thinking than is often supposed. At the same time, its clarity and conciseness suit it admirably for teaching purposes.

GREGORY CURRIE
University of Otago

REFERENCES

BENNETT, J. [1976]: *Linguistic Behaviour*. Cambridge University Press.
CURRIE, G. [1984]: 'Individualism and Global Supervenience', *British Journal for the Philosophy of Science*, **35**, pp. 345–58.
CURRIE, G. and ODDIE, G. [1980]: 'Changing Numbers', *Theoria*, **66**, pp. 148–64.
GRICE, H. P. [1957]: 'Meaning', *Philosophical Review*, **66**, pp. 377–88.
LEWIS, D. K. [1969]: *Convention*. Harvard University Press.
LEWIS, D. K. [1970]: 'How to Define Theoretical Terms', *Journal of Philosophy*, **67**, pp. 427–46.
LEWIS, D. K. [1972]: 'Psychophysical and Theoretical Identifications', *Australasian Journal of Philosophy*, **50**, pp. 249–58.
STALNAKER, R. C. [1967]: 'Events, Periods and Institutions in Historians' Language', *History and Theory*, **6**, pp. 159–79.

---

[1] For the argument see Lewis *ibid.*, pp. 154–6 and Bennett *ibid.*, p. 180. Grice's account of communication is developed in a series of papers beginning with Grice [1957].

OAKLANDER, L. NATHAN [1984]: *Temporal Relations and Temporal Becoming— A Defense of a Russellian Theory of Time*. University Press of America. xii + 238 pp.

Oaklander's book is an argument for a *B*-theory of time. As he sees it, the fundamental dispute between *A*-theories and *B*-theories concerns ontology—what is the proper ontological explanation of the true sentences that in ordinary language describe temporal relations (as when one event is said to be earlier or later than another) and temporal becoming (as when one event is said to have been future and become present). Not only is the *B*-theorist committed to denying that temporal becoming (*A*-facts) are 'the fundamental temporal realities', but, Oaklander argues, the *A*-theorist is equally committed to denying fundamental temporal reality to temporal relations (*B*-facts).

Like Hugh Mellor recently [1981], Oaklander adapts McTaggart's argument against the reality of time, to turn it into an argument against the reality of tense. He considers various *A*-theorist replies to this argument, and claims to answer them all. He is also like Mellor in that both argue that tensed statements are not reducible to tenseless statements, while denying that this entails the correctness of the *A*-theory. But unlike Mellor, Oaklander's defence of the *B*-theory very much depends on the success of his arguments against the *A*-theory, because whereas Mellor can justifiably claim to have produced a *positive* account of time that can stand on its own merits, Oaklander cannot claim the same intrinsic appeal for his account of time. Mellor's *B*-theory offers an explanation of the difference between time and space, the direction of time and its asymmetry, and the possibility and nature of change. Oaklander's *B*-theory offers much less. That time has a direction and is asymmetrical is, in effect, for Oaklander, a simple unanalysable fact about time. That time and space are different is, again, a primitive fact about time and space. Succession is what gives time its direction and its uniqueness—but succession is itself unanalysable and primitive. Oaklander doesn't have much in the way of a positive account of change, except that it 'consists in a variation of properties between successive events that constitute the history of an object' (p. 171). Disappointingly, however, he barely discusses rival *B*-theory positions, and doesn't mention Mellor's book at all (except in the preface). The inducements that Oaklander offers the reader for accepting his theory are his arguments purporting to show that the *A*-theory is contradictory, and some arguments purporting to answer some of the charges that are frequently brought against the *B*-theory, such as that it implies the correctness of fatalism. But his arguments in both of these areas are often contentious and unsatisfactory.

Oaklander sees the *A*-theory as being the view that reality embraces only that which presently exists. Part of his argument for the claim that the *A*-theory is committed to denying the fundamental temporal reality of *B*-facts is based on this assumption. Only at one point does he consider the alternative *A*-theory position (as defended by Broad) that reality embraces both present and past objects and events. But he dismisses it with the remark that this view requires that once an event has come into existence, it continues to exist thereafter—which, he says, cannot possible be true as Caesar is not now crossing the Rubicon. Certainly Broad would be in serious trouble if he had to say that Caesar *is* now crossing the Rubicon, but it's not at all obvious that he is committed to saying that. His position would at least look more plausible if it could be argued that there are two senses of 'exist'—in one of which it applies to anything that is part of reality (so that Aristotle exists in a way that Pegasus doesn't); and in the other sense it applies only to objects when they have the capacity to undergo change and have real properties, or to events, when they happen. If this were accepted, Broad could accept that past objects and events do not exist in the latter sense, but do exist in the former sense. In a later passage, Oaklander says that there *are* two senses of 'exist'. He needs this distinction to defend himself from the appearance of a contradiction as he wants to find a sense in which he can say both that all events exist, and that only the present exists. Admittedly, Oaklander's distinction is between tensed and tenseless existence, which is probably not quite the distinction that Broad needs. But argument is needed, and isn't provided, to show that Broad cannot make a similar distinction to sustain his view of temporal reality.

As well as seeing the *A*-theory as the position that only what is present is real, Oaklander also prefers to see the *A*-theory as being the view that past, present and future are properties of events. There are two good reasons (both of them pointed out to me by Nick Denyer) for thinking that this is a mistake. For one thing, it makes it extremely difficult to account for complex tenses. If we write '*e* will be past' as '*FPe*', of which event is '*F*' supposed to be predicated? *e*'s being past? '*Pe*' looks as though it should be some kind of truth-bearer, but it is difficult to see how the event of *e*'s being past (if it is one) could be a truth-bearer. The second point is that if we treated past, present and future as properties of events, how could we account for the possibility of *false* tensed statements? Of what event is pastness falsely being predicated in the sentence 'Julius Caesar kissed Queen Victoria'? Where a tensed statement is false, there will usually be no event of which pastness, presentness and futurity can, falsely, be predicated. Both of these points suggest that the approach of Prior and others in treating 'past', 'present' and 'future' not as predicates but as sentential operators is more natural and sensible. Given that, it requires much more attention than Oaklander gives it.

For Prior, all talk about the 'time-series' and 'instants of time' is disguised talk about what is and has been and will be the case. Oaklander charges Prior with presupposing that which he sets out to explain because, according to Prior, tensed statements are true at one time and false at another. And what, he asks, is this 'at one time' of which Prior speaks? Obviously the mere fact that Prior uses that phrase does not suffice to show him to be treating instants of time as primitive entities. It has to be shown that he *has* to be assigning primitive status to instants. But Oaklander does not show that. Oddly, he even admits later that it is not obvious to him that Prior has to refer to instants in specifying the truth-conditions for tensed statements. He says instead that Prior cannot treat 'it will be the case that' and 'it has been the case that' as operators without temporal significance, that Prior's tensed facts are *temporal* facts that exist in time, and that temporal operators are properties that such 'facts' possess. His reasons for saying this are that temporal operators are not truth-functional (he doesn't tell us why this matters) and that L. J. Cohen is right in claiming that the temporal truth-evaluations Prior uses in expressing the semantic rules for tensed statements, cannot be mere operators. But it's not clear why not.

Again and again Oaklander simply fails to treat the *A*-theories with the seriousness they deserve. He claims, for example, that Prior abandons the law of contradiction because Prior has argued that 'it will be that *p*' and 'it will be that not-*p*' can both be false. Oaklander simply assumes that from the falsity of 'it will be that not-*p*' we can deduce the truth of 'it will be that *p*'. Clearly, it is only this latter assumption that Prior needs to challenge, and not the law of contradiction. Whether or not Prior is right, he can't be quite so easily dismissed.

I turn finally to Oaklander's version of the McTaggart argument. This purports to show that treating 'past', 'present' and 'future' as either monadic predicates or as relations that hold between events and the NOW, leads to contradiction. The conclusion for Oaklander is that the *A*-theory is paradoxical and inconsistent.

The argument is the familiar one that past, present and future are incompatible properties of events. All events can have one and only one of these properties; but every event must have all three. The objection is then considered that the fact that all events have all three does not matter, because no event has all three at once. Rather, they are had successively. But, the argument goes, this simply reraises the same contradiction at another level because an event's being past (or present or future) is itself something that it has in the past, in the present and in the future. And, again, these are all incompatible. If the contradiction arises at the first level, it can't be removed by compounding tenses to move to higher levels. *If* the contradiction arises at the first level. But does it?

Instead of agreeing that all events are past, present and future and then attempting to 'remove the contradiction' (an odd idea when one thinks about it) the *A*-theorist can simply deny that all events *are* past, present and future. My first game of Chess is past. It is not present, nor is it future. Ergo, not all events are past, present and future. If the *A*-theorist takes this line, he has not (yet) contradicted himself. So, as yet, he has no case to answer. The sense in which my first game of Chess is past, and not present or future, is the most ordinary sense in which we use these terms. If there is another sense in which that event *is* present and future, the onus is now on Oaklander to find it, and show why this poses a problem for the A-theorist. That does not seem to be an easy thing to do. It is certainly true that my first game of Chess was once future, was once present and has been past for some time. But the movement to this level is not now a retreat from a contradiction. As no contradiction has yet been found at the first level, no contradiction can yet be shown to occur at the second level either. The A-theorist can simply deny that as well as having been future and having been present, it also is now future and is now present, or will be future or present. One might, perhaps, get the A-theorist to agree that for every event, there is some time at which it is past, some time at which it is future, and some time at which it is present. But there is nothing contradictory about that, and the A-theorist is not being *forced* to accept this as the reality underlying use of tensed statements; so it doesn't help Oaklander.

Although Oaklander considers several objections to his McTaggartian argument, he sees all of them as attempts to 'remove the contradiction' and to explain how events can be past, present and future. We can agree that once a *prima facie* contradiction has been conceded, all attempts to remove it are doomed to failure. With this argument, the battle must be won or lost in the first skirmish—the attempt to establish a *prima facie* contradiction. But Oaklander doesn't provide us with any sense in which all events are or must be past, present and future, on any view of time. He says only that that this is the case on the A-theory is a consequence of the fact that, on the A-theory, time involves passage. He explains this in McTaggart's terms by saying that this means 'the movement of *presentness* along the [static, fixed, 'nontemporal'] C-series' (p. 49). And this, I think, is the clue to the appeal of McTaggart's argument. The contradiction arises when one tries to superimpose two mutually incompatible views of time on top of each other, by seeing time as *both* static and dynamic. From this initial inconsistency, further inconsistencies are easily generated. If, on the A-theory, passage is the movement of presentness along the C-series, McTaggart's claim, cited approvingly by Oaklander, that the A-theory presupposes that which it sets out to explain, is clearly correct. But all that this shows is that this conception of the A-theory is hopeless. Oaklander needs to show that the A-theorist is committed to it. This he fails to do.

SIMON ORDE
Clare College, Cambridge

REFERENCES

MELLOR, D. H. [1981]: *Real Time*. Cambridge University Press.

ALEXANDER, PETER [1985]: *Ideas, Qualities and Corpuscles: Locke and Boyle on the External World.* Cambridge University Press. x+330 pp. £27.50.

In recent years quite a substantial number of books on Locke's *Essay* and its intellectual background have been published. *Ideas, Qualities and Corpuscles* is both narrower and broader than most of its competitors: narrower in that it is concerned with only a selection of the subjects covered in the *Essay*, but broader in that it includes a substantial discussion of Robert Boyle's thoughts on these same topics.

Some aspects of the book have a slightly old-fashioned air. For example, Alexander emphasizes in several places the dangers of using abridgements of the *Essay* rather than the whole work. It is not that one disagrees with this; it is merely that it seems surprising that Alexander thinks that the point needs to be made at all. Abridgements of the *Essay* are strictly for undergraduates in a hurry. In fact Alexander's book still seems based on a rather restricted range of Lockean sources: the *Essay* itself and the correspondence with Stillingfleet. There are almost no references to any of Locke's other works, and there is no use whatever of the manuscript material in the Lovelace Collection.

Ideas, Qualities and Corpuscles is divided into two parts: there are three chapters on Boyle in Part I, and eleven chapters on Locke in Part II. The first and third of the chapters on Boyle contain useful discussions of Boyle's views about method and about the corpuscularian philosophy. The second chapter (Boyle and the Peripatetics) is an attempt to go deeper into the historical background to Boyle's thought. It is in my opinion less successful, primarily because Alexander does not attempt to make any independent investigation into the views of the scholastic authors whom Boyle was attacking. They are seen purely through Boyle's eyes. (Here a contrast with the history of science is instructive. The time has long past when it was acceptable to reconstruct the views of Galileo's peripatetic opponents purely from Galileo's portrait of them.) The only other evidence which Alexander uses comes from Aristotle himself.

One result of this is that quite large parts of Alexander's analysis are in effect pure guesswork—intelligent guesswork, but guesswork nonetheless. For example, in discussing Boyle's objections to the scholastic doctrine of real qualities, he remarks that 'this conception of real qualities probably arose in the following way' (p. 49), after which there follows a long paragraph containing no references to any primary sources or to anything in the secondary literature. Not surprisingly, the account is quite inadequate: in particular, it makes no mention of the fact that for the scholastics who put forward this theory not all qualities are real qualities (*e.g.* Ockham, *Summa Logicae*, I.55; *Quodlibetae Septem* VII q. 7; *De Sacramento Altaris*, ch. 12).

This failure to make use of medieval and post-medieval scholastic material heads to other important misunderstandings. Alexander holds that

Aristotle used the theory of the four elements in his physics and the theory of matter and form in his metaphysics, and that it was only during the Middle Ages that the distinction between the two became blurred. This seems very doubtful: the concepts of form and matter are first introduced in *Physics* I and are ubiquitous thereafter, and in any case the idea that Aristotle's physical and metaphysical enquiries can be clearly separated hardly stands up to examination (where, for example, would one place *Metaphysics* Λ?). It does however lead Alexander to suggest that the schoolmen went wrong in building a physical theory involving causal explanations out of 'what for Aristotle was merely a logical analysis of change leading to the metaphysical assertion of prime matter' (p. 51). The objection to this is that it projects a fundamentally modern distinction between logical (or conceptual) and scientific (or factual) enquiries onto Aristotle, to whom the distinction is quite foreign. Aristotle's celebrated analysis of motion in the *Physics* is neither a piece of conceptual analysis nor an *a posteriori* enquiry into how bodies move. Boyle was already thinking in the modern way, and this is one reason why he found much of what Aristotle and the scholastics were trying to do to be quite incomprehensible.

Some of the same kinds of misunderstanding appear elsewhere. In chapter 4 Alexander is concerned with Locke's use of the word 'idea' and with Locke's explanation that he has used it 'to express whatever is meant by *Phantasm*, *Notion*, *Species*' (*Essay*, I.i.8). First he cites Kemp Smith as evidence that the Latin translation of both *ἰδέα* and *εἶδος* was *species* (p. 9). This is true for *εἶδος*, but false for *ἰδέα*. Admittedly Cicero translated *ἰδέα* by *species*, (*Academica*, i.30), but other writers (*e.g.* Seneca) used '*idea*'. The scholastics always referred to the Platonic Ideas as *ideae*. Alexander then goes on to say more about the scholastic doctrine of species, which he sees as descending not only from Aristotle but also from Democritus and Epicurus. The images which the atomists supposed to be emitted by bodies 'were called *εἶδωλα* and, in Latin, *species* or *effluvia*. Although Aristotle does not appear to hold a doctrine of effluvia many of his medieval followers did' (p. 98). This is also misleading. The usual translation of *εἶδωλον* is '*simulacrum*' (as in Lucretius), and those medieval theories of light which employ a notion of species (*e.g.* that found in Roger Bacon's *De Multiplicatione Specierum*) are utterly different from those of the atomists.

Most of the faults of Alexander's discussion of these topics arise from his failure to go back to the primary sources. All the remarks about the character of the various scholastic theories are taken from the notes to Sir William Hamilton's edition of Reid (1846). Hamilton was a man of wide learning and his footnotes contain much recondite information, but it is fair to say that as a guide to scholastic philosophy his work has been superseded.

The parts of the book which are concerned with the analysis of Locke's thought are more successful than those which deal with the historical background. Chapters 5–10 cover Locke's account of primary and secondary qualities. Here Alexander makes a number of closely related claims. One is that the distinction between primary and secondary qualities is grounded in the corpuscularian philosophy. It is not a 'philosophical' distinction (in the modern sense of that word) which can be detached from a particular scientific research programme. This seems to me to be entirely correct. Secondly, Locke's theory of primary qualities is generally misunderstood. There are only three primary qualities—size, shape, and mobility. The other so-called primary qualities are either secondary qualities, or else not qualities at all. Thirdly the secondary qualities of bodies *are* the textures of their invisible parts (usually counted among primary qualities). These are certainly substantial claims. Unfortunately they seem to me to be rather doubtful. Consider the case of solidity, which is usually considered as a primary quality, not least because Locke appears quite definitely to say so, as when he refers to 'the *primary*, and *real Qualities* of Bodies, which are always in them, (*viz*. Solidity, Extension . . .' (*Essay*, II. viii. 22). Alexander uses two arguments against the view that solidity is a primary quality. One is that qualities are what differentiate bodies from one another, and that since all the primary corpuscles are equally solid, solidity cannot be a quality, though it is essential to matter (pp. 139–40). The other argument is that a quality is defined by Locke as 'the Power to produce any *Idea* in our mind' (*Essay*, II. viii. 8), and that since the solidities of the corpuscles do not produce any *particular* idea in our minds, solidity is not a quality. Neither of these arguments is very strong. If Locke had explicitly denied that solidity is a primary quality then they might be acceptable as possible reconstructions of his line of thought. Since Locke does not say this, and appears to say the very opposite, both arguments would have to be much stronger in order to carry conviction.

There is much more in the book which deserves discussion. Indeed one of its great merits is that it does continually provoke thought about what Locke was trying to do. I have found myself unable to agree with many of Alexander's conclusions, but his arguments are never silly, or baseless, or merely perverse. Despite some faults of historical scholarship *Ideas, Qualities and Corpuscles* is likely to be of real value to students of seventeenth-century philosophy.

J. R. MILTON
Imperial College, London

COSMIC UNDERSTANDING

Philosophy and Science of the Universe

Milton K. Munitz

"This provocative book should be widely accessible to intelligent readers who are neither astrophysicists nor philosophers. It is the best 'popular' exposition of scientific cosmology that I have read for the purpose of conveying the limitations of current scientific knowledge of the 'beginnings of the Universe.' The author attempts to recognize a transcendent level of being (beyond, but not incompatible with, science) but without a religious component."
— [illegible]

Given the prevailing consensus among cosmologists that the universe had its beginning approximately 15 billion years ago, and that it will reach its end in the remote yet foreseeable future, we face the momentous intellectual challenge of how to assimilate these scientific claims into our fundamental world view. In this work the distinguished philosopher Milton Munitz provides a lucid account of the chief empirical findings and theories of recent cosmology and a systematic assessment of their broader philosophical implications.

$25.00 *At your bookstore or*

Princeton University Press

41 William Street, Princeton, NJ 08540

The Journal of Philosophy

SUBSCRIPTIONS

Individuals	$20.00
Libraries and institutions	$25.00
Students, retired/unemployed philosophers	$15.00
Postage outside U.S.	$ 5.00

BACK VOLUMES

Published bi-monthly 1904–1976, monthly thereafter.

All back volumes and individual issues are available, also a Cumulative Fifty-year Index, 1904–1953 and a Ten-year Supplemental Index, 1954–1963. Back volumes 1976–present also available in microform.

Please inquire for price lists and shipping charges on back orders.

720 PHILOSOPHY HALL, COLUMBIA UNIVERSITY, NYC 10027

The Monist

An International Quarterly Journal of General Philosophical Inquiry

Editor, John HOSPERS

Editorial Board: Henry Allison, William P. Alston, John M. Cooper, Dagfinn Føllesdal, R. Barcan Marcus, Joseph Margolis, Wallace I. Matson, Joseph Owens, Bruce Vermazen, Wilfrid Sellars, Anita Silvers, John E. Smith, Richard Wasserstrom.

Managing Editor: Sherwood J. B. SUGDEN

EACH ISSUE IS LIMITED TO ARTICLES ON A SINGLE GENERAL TOPIC. Communicate in advance with the Editor for special instructions. Two nonreturnable copies (one copy outside the U.S. and Canada) should be submitted NINE MONTHS PRIOR to the announced publication date. Papers should be prepared for anonymous submission to readers, and should be from 10 to 20 double-spaced typewritten pages in length, including notes.

SCHEDULED PUBLICATION DATES:

Vol. 71, No. 1	Jan.,	1988	Philosophical Problems of Space Exploration
Vol. 71, No. 2	Apr.,	1988	Aesthetics and the Histories of the Arts
Vol. 71, No. 3	July,	1988	Justification
Vol. 71, No. 4	Oct.,	1988	Descartes and His Contemporaries
Vol. 72, No. 1	Jan.,	1989	Logic as a Field of Knowledge
Vol. 72, No. 2	Apr.,	1989	Kant's Critical Philosophy
Vol. 72, No. 3	July,	1989	Kant's Practical Philosophy
Vol. 72, No. 4	Oct.,	1989	The Philosophy of R. G. Collingwood
Vol. 73, No. 1	Jan.,	1990	Hellenistic Ethics

Editorial Office: School of Philosophy, University of Southern California
Los Angeles, CA 90089
Business Office: Box 600, La Salle, Illinois 61301

SUBSCRIPTION RATES: United States: Annual (4 issues): Institutions, $24.00; Individuals, $15.00. Single copies: $6.00. Foreign postage: Add 50 cents to single copy rate or $1.00 to subscription.

ARE YOU STILL READING SOMEONE ELSE'S COPY OF ISIS?

IF SO, now is the time to enter your own subscription. Isis, the official journal of the History of Science Society, is the leading journal in the field.

Isis keeps over 4,000 subscribers in nearly fifty countries up to date on all developments in the history of science with articles, critiques, and documents and translations. Its notes and correspondence and news of the profession provide useful information to professionals, educators, scholars, and graduate students.

Lively essay reviews and over 200 book reviews a year cover every specialty in the history of science, technology, and medicine.

In addition to your four quarterly issues of Isis you will also receive:

- Membership in the History of Science Society.
- The annual *Critical Bibliography*, listing over 3,000 publications in the history of science, technology and medicine from the preceding year.
- The triennial *Guide to the History of Science*, containing a directory of members and information on scholarly programs, societies, and over ninety journals in the field. The 1983 edition is now available.
- The quarterly *Newsletter*, providing current news of the profession, including employment opportunities and approaching meetings.

ISIS

History of Science Society
Publications Office, Dept. GC
University of Pennsylvania
215 South 34th St./D6
Philadelphia, Pa. 19104

Or call TOLL-FREE
(orders only, please!)
1-800-341-1522
DATATEL-800™
6AM-9PM EST

GRAZER PHILOSOPHISCHE STUDIEN

BAND 27 | 1986 | VOLUME 27

Aufsätze | *Articles*

Some Clarifications and Replies . . . . Stephan KÖRNER
Frege and the Case of the Missing Sense . . . . Pavel TICHY
Plantinga on the *De Dicto/De Re* Distinction . . . . Michael WREEN
Thinking about Thinking . . . . Joseph MARGOLIS
Epistemic Coherentism and the Isolation Objection . . . . Paul K. MOSER
Cartesian Uncertainty: Descartes and Dummett . . . . M. GLOUBERMAN
Was sind Sinnesdaten? . . . . Heinz Dieter HECKMANN
Jean Cavaillès and the Vienna Circle . . . . Santiago RAMIREZ
Counting the Costs of Equality . . . . Kai NIELSEN

Besprechungen | *Review Articles*

Edmund HUSSERL: Logische Untersuchungen, II. Band, 1. und 2. Teilband (Husserliania XIX/1, XIX/2), ed.: Ursula Panzer, The Hague/Boston/Lancaster: Martinus Nijhoff 1984 . . . . K. MULLIGAN, B. SMITH
Jürgen MITTLESTRASS (Hrsg.): Enzyklopädie – Philosophie und Wissenschaftstheorie, Mannheim/Wien/Zürich: Bibliographisches Institut, Bd. 1: A–G (1980), Bd. 2: H–O (1984) . . . . Rudolf HALLER
Gilbert HOTTOIS: Le Signe et la Technique; la philosophie à l' épreuve de la technique, Paris: Aubier 1984 . . . . J.-Y. GOFFI

Herausgeber | *Editor*

Prof. Dr. Rudolf Haller, Institut für Philosophie, Universität Graz, Heinrichstraße 26, A-8010 Graz, Österreich/Austria.

AUSTRALASIAN JOURNAL OF PHILOSOPHY

Contents of Vol. 65, No. 1; March 1987

EDITOR: BRIAN ELLIS

ARTICLES

Frederick W. Kroon .... Causal Descriptivism
Robert J. Lipkin .... Altruism and Sympathy in Hume's Ethics
George Schlesinger .... Accommodation and Prediction
Raymond D. Bradley .... Wittgenstein's Tractatarian Essentialism
Jordan Howard Sobel .... Self-Doubts and Dutch Strategies
Raymond Martin .. Memory, Connecting and What Matters in Survival
Gerald H. Paske ... Sperm-Napping and the Right Not to Have a Child

DISCUSSION

B.J. Copeland .... Appraising Historical Accounts: A Discussion of Gorman's View

BOOKS RECEIVED, NOTES AND NEWS, INDEX

ANNUAL SUBSCRIPTION: Individuals A$20 (US$20 or Stg 10); Institutions A$40, (US$40 or Stg 20). All correspondence should be addressed to the Editor or Business Manager, Department of Philosophy, La Trobe University, Bundoora, Victoria, 3083, Australia.

PUBLISHED QUARTERLY BY
THE AUSTRALASIAN ASSOCIATION OF PHILOSOPHY

WOMEN AND Philosophy
Special Issue

AUSTRALASIAN JOURNAL OF PHILOSOPHY

Edited by Janna Thompson with the assistance of Mary McCloskey Brenda Judge San MacColl Elizabeth Gross

ARTICLES

Moira Gatens .... Rousseau and Wollstonecraft: Nature vs Reason
Gareth B. Matthews .... Gender and Essence in Aristotle
Karen Green .... Rawls, Women and the Priority of Liberty
Marion Tapper .... Can a Feminist be a Liberal?
Lorraine Code .... Simple Equality Is Not Enough
John Howes .... Mill on Women and Human Development
Susan Parsons .... Feminism and Moral Reasoning
Jan Crosthwaite and Christine Swanton ...... On the Nature of Sexual Harassment
Robert Pargetter and Elizabeth W. Prior ..... Against the Sexuality of Reason

CRITICAL REVIEW

Val Plumwood .... Ecofeminism: An Overview and Discussion of Positions and Arguments

PRICE: A$15. All correspondence should be addressed to the Editor or Business Manager, Department of Philosophy, La Trobe University, Bundoora, Victoria, 3083, Australia.

PUBLISHED BY THE AUSTRALASIAN ASSOCIATION OF PHILOSOPHY